Online Appendices for

"The Accuracy of Aggregate Student Growth Percentiles as Indicators of Educator Performance"

Article published in *Educational Measurement: Issues and Practice*

doi: 10.1111/emip.12144

Katherine E. Castellano

Daniel F. McCaffrey

Educational Testing Service

A Derivations for Paper Results

This appendix provides the derivations for the results presented throughout this paper. See the paper for notation and distributional assumptions. The derivations are given for the mean aggregated-student estimators, such as the mean of the Standard SGP, in Section A.1 and the median aggregated-student estimators, such as the median of the Standard SGP, in Section A.2. To distinguish between mean and median estimators, we use meanGP and medGP instead of MGP to generically represent a mean or median SGP estimator, and to represent a specific estimator, we use meanEst, such as meanStd for the Standard mean estimator, and medEst, such as medStd for the Standard median estimator. We also dispense with the shorthand of $\text{EMGP}_{\text{Est}|\theta,\delta}$ and $\text{EMGP}_{\text{Est}|\theta}$, and instead refer to them as $E[\text{meanEst} \mid \theta, \delta]$ and $E[\text{meanEst} \mid \theta]$, respectively. The formulas in this section apply to aggregates of the Standard, True, SIMEX and Ranked SIMEX SGP that are included in the paper as well as to aggregates for the other alternative (aggregated) SGP estimators described in Appendix B (i.e., Double-Corrected, MLE, and EAP).

A.1 Mean-Aggregated SGP

A.1.1 $E[\text{meanGP} \mid \theta, \delta]$

To derive $E[\text{meanGP} \mid \theta, \delta]$, we fist note that $E[\text{meanGP} \mid \theta, \delta] = E[\text{SGP} \mid \theta, \delta]$ following standard results for the expected value of a sample mean of a random variable. Next, we note that, under our normality assumptions, after conditioning on θ and δ, each of the student-level SGP estimators of interest can be expressed as the normal CDF evaluated at a random variable of the form $\frac{\xi - a}{b}$ (so $\Phi\left(\frac{\xi-a}{b}\right)$), where $\xi \sim N(0, \sigma_\xi^2)$ is a random variable that represents all the random terms that remain after conditioning on θ and δ, and a represents the fixed terms (i.e., those that include θ and δ) in the numerator, while b is also a constant–the denominator of the student-level SGP. We can then simply apply an established result for normal distributions (e.g., Zacks, 1981, pp. 53-54):

$$E\left[\Phi\left(\frac{\xi - a}{b}\right)\right] = \Phi\left(\frac{-a}{\sqrt{b^2 + \sigma_\xi^2}}\right). \tag{A.1}$$

For instance, we obtain the result given in Equation 4 in the paper for the Standard MGP as follows:

$$\begin{aligned}
E[\text{meanStd} \mid \theta,\delta] &= E\left[\Phi\left(\frac{W_2-\lambda_1\beta W_1}{\sqrt{\sigma^2_{W_2}(1-\rho^2)}}\right)\middle|\theta,\delta\right] \\
&= E\left[\Phi\left(\frac{[\eta+U_2+\beta(1-\lambda_1)\zeta-\lambda_1\beta U_1]-[-\theta-(1-\lambda_1)\beta\delta]}{\sqrt{\sigma^2_{W_2}(1-\rho^2)}}\right)\right] \\
&= \Phi\left(\frac{\theta+(1-\lambda_1)\beta\delta}{\sqrt{\left[\sigma^2_{W_2}(1-\rho^2)\right]+\left[\sigma^2_\eta+\sigma^2_{U_2}+(1-\lambda_1)^2\beta^2\sigma^2_\zeta+\lambda_1^2\beta^2\sigma^2_{U_1}\right]}}\right) \\
&= \Phi\left(\frac{\theta+(1-\lambda_1)\beta\delta}{\sqrt{\left[\sigma^2_{W_2}(1-\rho^2)\right]+\nu_1}}\right), \qquad \text{(A.2)}
\end{aligned}$$

where $\xi = \eta + U_2 + \beta(1-\lambda_1)\zeta - \lambda_1\beta U_1$, $a = -\theta - (1-\lambda_1)\beta\delta$, $b^2 = \sigma^2_{W_2}(1-\rho^2)$, and $\sigma^2_\xi = \sigma^2_\eta + \sigma^2_{U_2} + (1-\lambda_1)^2\beta^2\sigma^2_\zeta + \lambda_1^2\beta^2\sigma^2_{U_1} = \nu_1$.

A.1.2 $E[\text{meanGP} \mid \theta]$

To obtain the target $E[\text{meanGP} \mid \theta]$ for each meanGP estimator, we can apply the same result given in Equation A.1 because $E[\text{meanGP} \mid \theta] = E\left[E[\text{meanGP} \mid \theta,\delta]|\theta\right]$ can also be expressed as the expected value of a random variable of the form $\Phi\left(\frac{\xi-a}{b}\right)$. For instance, after conditioning on θ, the numerator of $E[\text{meanStd} \mid \theta,\delta]$ can be decomposed into a fixed component, simply $a = -\theta$, and a random component, $\xi = (1-\lambda_1)\beta\delta$, which is normally distributed with mean 0 and variance $\nu_2 = (1-\lambda_1)^2\beta^2\sigma^2_\delta$, and the denominator of $E[\text{meanStd} \mid \theta,\delta]$ equals $b^2 = \sigma^2_{W_2}(1-\rho^2)+\nu_1$. Thus, we obtain:

$$E[\text{meanStd} \mid \theta] = \Phi\left(\frac{\theta}{\sqrt{\left[\sigma^2_{W_2}(1-\rho^2)\right]+\nu_1+\nu_2}}\right),$$

as shown in Equation 5.

A.1.3 $\text{var}\left(E[\text{meanGP} \mid \theta, \delta]|\theta\right)$

This variance is discussed in reference to Figure 2(b) in the paper. We first decompose it into terms that we can easily obtain using standard normal distribution results:

$$\begin{aligned}
Var(E[\text{meanGP} \mid \theta,\delta] \mid \theta) &= E\left[(E[\text{meanGP} \mid \theta,\delta] - E[\text{meanGP} \mid \theta])^2\middle|\theta\right] \\
&= E\left[E[\text{meanGP} \mid \theta,\delta]^2\middle|\theta\right] - 2E\left[E[\text{meanGP} \mid \theta,\delta]|\theta\right] E[\text{meanGP} \mid \theta] + E\left[E[\text{meanGP} \mid \theta]^2\middle|\theta\right] \\
&= E\left[E[\text{meanGP} \mid \theta,\delta]^2\middle|\theta\right] - E[\text{meanGP} \mid \theta]^2. \qquad \text{(A.3)}
\end{aligned}$$

The second term, $E[\text{meanGP} \mid \theta]^2$, is simply the square of the term derived in Section A.1.2 above. To obtain the first term, $E\left[E[\text{meanGP} \mid \theta,\delta]^2|\theta\right]$, we note that after conditioning on θ, $E[\text{meanGP} \mid \theta,\delta]$ can be expressed as a random variable of the form $\Phi\left(\frac{\xi-a}{b}\right)$, and then use the result for the expectation of a square of the normal CDF evaluated at a normal random variable (see e.g., Zacks, 1981, pp. 53-54). That is,

$$E\left[\Phi\left(\frac{\xi-a}{b}\right)^2\right] = Pr\left(Z_1 \leq \frac{\xi-a}{b}, Z_2 \leq \frac{\xi-a}{b}\right) = Pr(t_1 \leq 0, t_2 \leq 0), \qquad \text{(A.4)}$$

where Z_1, Z_2 are independent, standard normal variables and t_1, t_2 are bivariate normal with mean vector (a, a), variances $b^2 + \sigma_\xi^2$, covariance σ_ξ^2. For example, for the meanStd, $E\left[E[\text{meanStd} \mid \theta,\delta]^2|\theta\right]$ is the bivarate normal probability with mean vector $(-\theta,-\theta)$, variances $\sigma_{W_2}^2(1-\rho^2) + \nu_1 + \nu_2$, and covariance ν_2.

A.1.4 $E\left[\mathrm{var}\left(\mathrm{meanGP} \mid \theta, \delta\right) \mid \theta\right]$

This expected value of a variance is discussed in reference to Figure 2(a) in the paper. We can decompose this term as follows:

$$\begin{aligned}
E\left[Var(\mathrm{meanGP}|\theta,\delta)|\theta\right] &= Var(\mathrm{meanGP}|\theta) - Var\left(E[\mathrm{meanGP}|\theta,\delta]|\theta\right) \\
&= \left(E[\mathrm{meanGP}^2|\theta] - E[\mathrm{meanGP}|\theta]^2\right) - \left(E\left[E[\mathrm{MGP}|\theta,\delta]^2|\theta\right] - E\left[E[\mathrm{meanGP}|\theta,\delta]|\theta\right]^2\right) \\
&= E[\mathrm{meanGP}^2|\theta] - E\left[E[\mathrm{meanGP}|\theta,\delta]^2|\theta\right] \\
&= \left(\frac{1}{n}E[\mathrm{SGP}^2|\theta] + \frac{n-1}{n}E\left[E[\mathrm{SGP}|\theta,\delta]^2|\theta\right]\right) - E\left[E[\mathrm{meanGP}|\theta,\delta]^2|\theta\right] \\
&= \left(\frac{1}{n}E[\mathrm{SGP}^2|\theta] + \frac{n-1}{n}E\left[E[\mathrm{meanGP}|\theta,\delta]^2|\theta\right]\right) - E\left[E[\mathrm{meanGP}|\theta,\delta]^2|\theta\right] \\
&= \frac{1}{n}\left(E[\mathrm{SGP}^2|\theta] - E\left[E[\mathrm{meanGP}|\theta,\delta]^2|\theta\right]\right), \qquad \text{(A.5)}
\end{aligned}$$

as noted in the paper. Both terms within the parentheses are expected values of the square of the normal CDF evaluated at normal random variables that can be found by applying the result given in Equation A.4. For instance, for the Standard SGP, the first term is the bivariate normal probability $E[\mathrm{SGP_{Std}}^2|\theta] = Pr[t_1 \le 0, t_2 \le 0]$, where t_1, t_2 are bivariate normal with mean vector $(-\theta, -\theta)$, variances $\left[\sigma^2_{W_2}(1-\rho^2)\right] + \nu_1 + \nu_2$ and covariance $\nu_1 + \nu_2$. Similarly, the second term is the bivariate normal probability: $E\left[E[\mathrm{meanStd}|\theta,\delta]^2|\theta\right] = Pr[t_1 \le 0, t_2 \le 0]$, where mean vector $(-\theta, -\theta)$, variances $\left[\sigma^2_{W_2}(1-\rho^2) + \nu_1\right] + \nu_2$ and covariance ν_2. Thus, the only difference between these two bivariate normal probabilities is their covariance, which corresponds to the variance of the random component after conditioning on θ. That is, for $E[\mathrm{SGP_{Std}}^2|\theta]$, $\xi = \eta + U_2 + \beta(1-\lambda_1)\zeta - \lambda_1\beta U_1 + (1-\lambda_1)\beta\delta$ and thus $\sigma^2_\xi = \nu_1 + \nu_2$, whereas for $E\left[E[\mathrm{meanStd}|\theta,\delta]^2|\theta\right]$, $\xi = (1-\lambda_1)\beta\delta$ and $\sigma^2_\xi = \nu_2$.

A.1.5 $E\left[\mathrm{var}\left(\mathrm{meanGP} \mid \theta\right)\right]$ or MSE

In Equation 8, we introduced $E\left[\mathrm{var}\left(\mathrm{meanGP} \mid \theta\right)\right]$ as the the error variance, or mean square error, treating $E[\mathrm{meanGP} \mid \theta]$ as the target for our meanGP estimator of interest.

We can derive var (meanGP | θ) as follows:

$$\text{var}(\text{meanGP} \mid \theta) = \frac{1}{n}\text{var}(\text{SGP} \mid \theta) = \frac{1}{n}\left[E(\text{SGP}^2 \mid \theta) - E(\text{SGP} \mid \theta)^2\right], \tag{A.6}$$

where each of the terms within the square brackets can be found using our general results for the normal distribution. For instance, for meanStd, both of these terms have already been derived in Sections A.1.2 and A.1.4.

To derive the expected value of this variance, however, we do not use the formula of the variance given in Equation A.6, instead we do the following:

$$\begin{aligned}
E\left[\text{var}\left(\text{meanGP} \mid \theta\right)\right] &= E\left[E\left[(\text{meanGP} - E[\text{meanGP} \mid \theta])^2 \middle| \theta\right]\right] \\
&= E\left[E\left(\text{var}[\text{meanGP} \mid \theta, \delta] \middle| \theta\right)\right] + E\left[\text{var}(E[\text{meanGP} \mid \theta, \delta] \middle| \theta)\right] \\
&= \frac{1}{n}\left(E[\text{SGP}^2] - E\left[E[\text{meanGP}|\theta, \delta]^2\right]\right) \\
&+ \left(E\left[E[\text{meanGP} \mid \theta, \delta]^2\right] - E\left[E[\text{meanGP} \mid \theta]^2\right]\right) \\
&= \frac{1}{n}E[\text{SGP}^2] + \frac{n-1}{n}E\left[E[\text{meanGP}|\theta, \delta]^2\right] - E\left[E[\text{meanGP} \mid \theta]^2\right], \quad \text{(A.7)}
\end{aligned}$$

where each of the three distinct terms equals the expected value of the square of a normal CDF evaluated at a normal random variable and can thus be found by applying Equation A.4. For instance, for the meanStd, each of these are bivariate normal probabilities $Pr[t_1 \leq 0, t_2 \leq 0]$, where t_1 and t_2 are bivariate normal with the following mean and variance structures:

1. $E[\text{SGP}_{\text{Std}}{}^2]$: mean vector $(0, 0)$, variances $\sigma^2_{W_2}(1 - \rho^2) + \nu_1 + \nu_2 + \nu_3$ and covariance $\nu_1 + \nu_2 + \nu_3$

2. $E\left[E[\text{meanStd}|\theta, \delta]^2\right]$: mean vector $(0, 0)$, variances $\sigma^2_{W_2}(1 - \rho^2) + \nu_1 + \nu_2 + \nu_3$ and covariance $\nu_2 + \nu_3$

3. $E\left[E[\text{meanStd} \mid \theta]^2\right]$: mean vector $(0, 0)$, variances $\sigma^2_{W_2}(1 - \rho^2) + \nu_1 + \nu_2 + \nu_3$ and covariance ν_3,

where $\nu_1 = \sigma^2_\eta + \sigma^2_{U_2} + (1 - \lambda_1)^2\beta^2\sigma^2_\zeta + \lambda_1^2\beta^2\sigma^2_{U_1}$, $\nu_2 = (1 - \lambda_1)^2\beta^2\sigma^2_\delta$, and $\nu_3 = \sigma^2_\theta$. Thus, these terms only differ in the covariances of the bivariate normal random variables, as the covariances correspond to the variance of the remaining random terms in the numerators of each of the squared normal CDFs: $E[\text{SGP}_{\text{Std}}], E[\text{meanStd}|\theta, \delta]$, and $E[\text{meanStd}|\theta]$.

A.1.6 var $(E[\text{meanGP} \mid \theta])$

The var $(E[\text{meanGP} \mid \theta])$, as noted in the paper, is independent of measurement error. It represents the variance in true class differences, or the variance of the target and is used in the proportional reduction in MSE formula (see Equation 9 in the paper). This variance can be derived using the general results for normal distributions as given in Equations A.1 and A.4:

$$Var[E(\text{meanGP} \mid \theta)] = E\left[E(\text{meanGP} \mid \theta)^2\right] - E\left[E(\text{meanGP} \mid \theta)\right]^2. \tag{A.8}$$

For instance, for Standard meanGP, when not conditioning on any terms, the $E(\text{meanStd} \mid \theta)$ can be written in our desired form, $\Phi\left(\frac{\xi - a}{b}\right)$, with $\xi = \theta$, $a = 0$ (i.e., there are no fixed terms), $b^2 = \sigma^2_{W_2}(1 - \rho^2) + \nu_1 + \nu_2$, and $\sigma^2_\xi = \sigma^2_\theta = \nu_3$. Thus, the first term in the above equation follows from Equation A.4 and is a bivariate normal probability $Pr[t_1 \leq 0, t_2 \leq 0]$, where t_1, t_2 are bivariate normal with mean vector $(0, 0)$, variances $\sigma^2_{W_2}(1 - \rho^2) + \nu_1 + \nu_2 + \nu_3$ and covariance ν_3. The second term reduces to the evaluation of the normal CDF or univariate normal probability following the result in Equation A.1: $E\left[E(\text{meanStd} \mid \theta)\right] = \Phi\left(\frac{0}{\sqrt{\sigma^2_{W_2}(1-\rho^2)+\nu_1+\nu_2+\nu_3}}\right) = \Phi(0) = .5$. That is, the unconditional expected value of the target is an meanGP of 50.

A.2 Median-Aggregated SGP

In the paper, we only present findings for mean aggregated SGP, but we could also consider the median aggregated SGP, which was first proposed by Betebenner (2009) as SGPs are only ordinal, not interval-scaled, measures. The results for the median aggregated student SGP all use the asymptotic distribution and thus are approximations which generally are accurate for large class sizes $n = 50$ or more. For clarity of presentation, the text does not repeatedly refer to the values as approximations but all results should be considered as such.

A.2.1 $E[\text{medGP} \mid \theta, \delta]$

We obtain the asymptotic expected value of a medGP given θ and δ by well-established statistical results for medians. That is, as the number of students in a class increases (goes to infinity), the asymptotic expected value of each medGP given both θ and δ is the population median.

As in Section A.1.1, conditional on θ and δ, we re-express the student SGP estimators in the form $\Phi\left(\frac{\xi-a}{b}\right)$, where given θ, ξ is random ($N(0, \sigma_\xi^2)$) and a and b are constants. Then, using established results for the normal distribution, the asymptotic expected median SGP is:

$$E[\text{medGP} \mid \theta, \delta] \approx \Phi\left(\frac{-a}{b}\right). \tag{A.9}$$

So, for instance, for the Standard SGP, as we have previously defined in Section A.1.1, conditional on θ and δ, the Standard SGP can be written in the desired form of the normal CDF evaluated at $\frac{\xi-a}{b}$, $\xi = \eta + U_2 + \beta(1-\lambda_1)\zeta - \lambda_1\beta U_1$, $a = -\theta - (1-\lambda_1)\beta\delta$, $b^2 = \sigma_{W_2}^2(1-\rho^2)$, and $\sigma_\xi^2 = \nu_1$. Accordingly,

$$E[\text{medStd} \mid \theta, \delta] \approx \Phi\left(\frac{\theta + (1-\lambda_1)\beta\delta}{\sqrt{\sigma_{W_2}^2(1-\rho^2)}}\right), \tag{A.10}$$

which corresponds to the student SGP estimator evaluated at the population mean of possible numerator values for the teacher ($\theta + (1-\lambda_1)\beta\delta$), and note that, unlike for the expected value of the meanGP estimators, the denominator contains no additional terms than those in the student-level SGP.

A.2.2 $E[\text{medGP} \mid \theta]$

The target $E[\text{medGP} \mid \theta]$ equals $E\left[E[\text{medGP} \mid \theta, \delta] | \theta\right]$. Thus because $E[\text{medGP} \mid \theta, \delta]$ (Equation A.9) equals the normal CDF evaluated at a normal random variable, the result given in Equation A.1 is applicable. For instance, for the medStd, conditional on θ, $E[\text{medStd} \mid \theta, \delta]$ can be expressed in our desired form, $\Phi\left(\frac{\xi-a}{b}\right)$: as the numerator ($\theta + (1-\lambda_1)\beta\delta$) given in Equation A.10 can be decomposed into a random component $\xi = (1-\lambda_1)\beta\delta \sim N(0, \nu_2)$ and a fixed component

$a = -\theta$. Thus:

$$E[\text{medStd} \mid \theta] = E\left[E[\text{medStd} \mid \theta, \delta] | \theta\right] = \Phi\left(\frac{\theta}{\sqrt{\sigma^2_{W_2}(1-\rho^2) + \nu_2}}\right). \tag{A.11}$$

A.2.3 $\text{var}\left(E[\text{medGP} \mid \theta, \delta] | \theta\right)$

The $\text{var}\left(E[\text{medGP} \mid \theta, \delta] | \theta\right)$ can be found by substituting meanGP with medGP in Equation A.3:

$$Var(E[\text{medGP} \mid \theta, \delta] \mid \theta) = E\left[E[\text{medGP} \mid \theta, \delta]^2 | \theta\right] - E[\text{medGP} \mid \theta]^2. \tag{A.12}$$

The second term was given in Section A.2.2, and the first term follows from the result given in Equation A.4. That is, as $E[\text{medGP} \mid \theta, \delta]$ is the normal CDF evaluated at a normal random variable, the expectation of this term squared given θ is the expectation of the square of the normal CDF evaluated at a normal random variable and is thus a bivariate normal probability. For instance, for the medStd, $E\left[E[\text{medStd} \mid \theta, \delta]^2 | \theta\right] = Pr(t_1 \leq 0, t_2 \leq 0)$ where t_1, t_2 are bivariate normal with mean vector $(-\theta, -\theta)$, variances $\sigma^2_{W_2}(1-\rho^2) + \nu_2$, and covariance ν_2.

A.2.4 $E\left[\text{var}\left(\text{medGP} \mid \theta, \delta\right) \mid \theta\right]$

The $E\left[\text{var}\left(\text{medGP} \mid \theta, \delta\right) \mid \theta\right]$ involves taking the expected value of the (asymptotic) variance, $\text{var}\left(\text{medGP} \mid \theta, \delta\right)$. So, we first derive this variance using common asymptotics for sample medians:

$$\text{Var}(\text{medGP} \mid \theta, \delta) \approx \frac{.5^2}{n f_{SGP|\theta,\delta}(E[\text{medGP} \mid \theta, \delta])^2}, \tag{A.13}$$

where $f_{SGP|\theta,\delta}(E[\text{medGP} \mid \theta, \delta])$ is the density function of the student-level SGP given θ and δ evaluated at $E[\text{medGP} \mid \theta, \delta]$. The density of the normal CDF evaluated at a normal random variable also follows from established results for normal distributions. Again, expressing a student-level SGP estimator in the desired form of $\Phi\left(\frac{\xi - a}{b}\right)$ after conditioning on θ and δ, it can be shown that the CDF of an SGP estimator is

$$F_{SGP|\theta,\delta}(k) = Pr\left[\Phi\left(\frac{\xi - a}{b}\right) \leq k\right] = Pr\left[\Phi^{-1}\left(\Phi\left(\frac{\xi - a}{b}\right)\right) \leq \Phi^{-1}(k)\right] = \Phi\left(\frac{b\Phi^{-1}(k) + a}{\sigma_\xi}\right)$$

Accordingly, the pdf of an SGP estimator is:

$$\begin{aligned} f_{SGP|\theta,\delta}(k) &= \frac{\partial}{\partial_k}\Phi\left(\frac{b\Phi^{-1}(k) + a}{\sigma_\xi}\right) \\ &= \phi\left(\frac{b\Phi^{-1}(k) + a}{\sigma_\xi}\right)\left(\frac{b}{\sigma_\xi}\right)\frac{\partial}{\partial_k}\Phi^{-1}(k) \\ &= \left(\frac{b}{\sigma_\xi}\right)\frac{\phi\left(\frac{b\Phi^{-1}(k)+a}{\sigma_\xi}\right)}{\phi(\Phi^{-1}(k))}. \end{aligned} \tag{A.14}$$

In our case, $k = E[\text{medGP} \mid \theta, \delta] \approx \Phi\left(\frac{-a}{b}\right)$ (Equation A.9). Thus, substituting this term in for k, Equation A.14 simplifies further as $\Phi^{-1}(k) = \Phi^{-1}(E[\text{medGP} \mid \theta, \delta]) = \Phi^{-1}\left[\Phi(\frac{-a}{b})\right] = \frac{-a}{b}$. Thus,

$$f_{SGP|\theta,\delta}(E[\text{medGP} \mid \theta, \delta]) = \left(\frac{b}{\sigma_\xi}\right)\frac{\phi(0)}{\phi(\frac{-a}{b})} = \left(\frac{b}{\sqrt{2\pi}\sigma_\xi}\right)\frac{1}{\phi\left(\frac{-a}{b}\right)}, \tag{A.15}$$

as $\phi(0) = \frac{1}{\sqrt{2\pi}}$.

Substituting this result for $f_{SGP|\theta,\delta}(E[\text{medGP} \mid \theta, \delta])$ in Equation A.13, we obtain:

$$\text{Var}(\text{medGP} \mid \theta, \delta) \approx \left(\frac{.5^2(2\pi)}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)\phi\left(\frac{-a}{b}\right)^2, \tag{A.16}$$

For instance, for Standard SGP, $a = -\theta - (1 - \lambda_1)\beta\delta$, $b^2 = \sigma_{W_2}^2(1 - \rho^2)$, and $\sigma_\xi^2 = \nu_1$ so we have:

$$\text{Var}(\text{medStd} \mid \theta, \delta) \approx \left(\frac{.5^2(2\pi)}{n}\right)\left(\frac{\nu_1}{\sigma_{W_2}^2(1 - \rho^2)}\right)\phi\left(\frac{\theta + (1 - \lambda_1)\beta\delta}{\sqrt{\sigma_{W_2}^2(1 - \rho^2)}}\right)^2.$$

We can now use this result to derive the expected value of this asympotitic variance given θ:

$$
\begin{aligned}
E\left[Var(\text{medGP} \mid \theta, \delta) \mid \theta\right] &\approx E\left[\left(\frac{.5^2\sigma_\xi^2(2\pi)}{nb^2}\right)\phi\left(\frac{-a}{b}\right)^2 \middle| \theta\right] \\
&= \left(\frac{.5^2(2\pi)}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)E\left[\phi\left(\frac{-a}{b}\right)^2 \middle| \theta\right] \\
&= \left(\frac{.5^2(2\pi)}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)E\left[\phi\left(\frac{X-\alpha}{b}\right)^2\right] \\
&= \left(\frac{.5^2(2\pi)}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)E\left[\frac{1}{2\pi}exp\left(\frac{-(X-\alpha)^2}{b^2}\right)\right] \\
&= \left(\frac{.5^2}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)\int \frac{1}{\sqrt{2\pi\sigma_X^2}}exp\left(\frac{-(X-\alpha)^2}{b^2} + \frac{-X^2}{2\sigma_X^2}\right)dX \\
&= \left(\frac{.5^2}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)\int \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\frac{\sigma_X^2}{1+2\sigma_X^2/b^2}}}exp\left(\frac{-\left(X - \frac{2\alpha\sigma_X^2}{b^2(1-2\sigma_X^2/b^2)}\right)^2}{2\sigma_X^2/(1+2\sigma_X^2/b^2)}\right)dX \\
&\quad \sqrt{\frac{1}{1+\frac{2\sigma_X^2}{b^2}}}exp\left(\frac{\left(\frac{2\alpha\sigma_X^2}{b^2(1+2\sigma_X^2/b^2)}\right)^2}{2\sigma_X^2/(1+2\sigma_X^2/b^2)} + \frac{\alpha^2}{b^2}\right) \\
&= \left(\frac{.5^2}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)\left(\frac{1}{\sqrt{1+\frac{2\sigma_X^2}{b^2}}}\right)exp\left(\frac{-\alpha^2}{b^2+2\sigma_X^2}\right). \qquad \text{(A.17)}
\end{aligned}
$$

Note that we re-express $\phi\left(\frac{-a}{b}\right)$ as $\phi\left(\frac{-X-\alpha}{b}\right)$ in line 3 as conditional on θ, the numerator $-a$ can be further decomposed into a random component X minus a fixed term α. The final result follows by noting that the expression within the integral is a normal pdf with variance $\frac{\sigma_X^2}{1+2\sigma_X^2/b^2}$ and thus integrating it over $-\inf$ to $+\inf$ equals 1.

Given there are several terms here, we review what this result is for the medStd. As noted previously, conditional on θ and δ, the Standard SGP can be expressed as $\Phi\left(\frac{\xi-a}{b}\right)$, where $\xi = \eta + U_2 + \beta(1-\lambda_1)\zeta - \lambda_1\beta U_1$, $a = -(\theta + (1-\lambda_1)\beta\delta)$, $b^2 = \sigma_{W_2}^2(1-\rho^2)$, and $\sigma_\xi^2 = \nu_2$. Thus, conditional on θ, we can decompose a into a fixed component $\alpha = -\theta$ and a random component $X = (1-\lambda_1)\beta\delta$ with $\sigma_X^2 = \nu_2$.

A.2.5 $E\left[\text{var}\left(\text{medGP} \mid \theta\right)\right]$ or MSE

As for the meanGP shown in Section A.1.5, the (asymptotic) $E\left[\text{var}\left(\text{medGP} \mid \theta\right)\right]$ equals the MSE for a medGP estimator with $E[\text{medGP} \mid \theta]$ as the target.

$$\begin{aligned}
E\left[\text{var}\left(\text{medGP} \mid \theta\right)\right] &= E\left[E\left[(\text{medGP} - E[\text{medGP} \mid \theta])^2 \middle| \theta\right]\right] \\
&= E\left[E\left(Var[\text{medGP} \mid \theta, \delta] \middle| \theta\right)\right] + E\left[Var(E[\text{medGP} \mid \theta, \delta] \middle| \theta)\right] \\
&= \left(E\left[Var[\text{medGP} \mid \theta, \delta]\right]\right) + \left(E\left[E[\text{medGP} \mid \theta, \delta]^2\right] - E\left[E[\text{medGP} \mid \theta]^2\right]\right)
\end{aligned} \tag{A.18}$$

The term in the first parentheses is the (asymptotic) unconditional expected value of $Var[\text{medGP} \mid \theta, \delta]$ and can be found using an derivations similar to those of the $E\left[Var(\text{medGP} \mid \theta, \delta) \mid \theta\right]$ given in Equation A.17, but the derivations are simpler here, as we are not conditioning on θ:

$$\begin{aligned}
E\left[Var(\text{medGP} \mid \theta, \delta)\right] &\approx E\left[\left(\frac{.5^2 \sigma_\xi^2 (2\pi)}{nb^2}\right) \phi\left(\frac{-a}{b}\right)^2\right] \\
&= \left(\frac{.5^2 (2\pi)}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right) E\left[\phi\left(\frac{-a}{b}\right)^2\right] \\
&= \left(\frac{.5^2}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right) \int \frac{1}{\sqrt{2\pi\sigma_a^2}} exp\left(\frac{-a^2}{b^2} + \frac{-a^2}{2\sigma_a^2}\right) da \\
&= \left(\frac{.5^2}{n}\right)\left(\frac{\sigma_\xi^2}{b^2}\right)\left(\frac{1}{\sqrt{1 + \frac{2\sigma_a^2}{b^2}}}\right).
\end{aligned} \tag{A.19}$$

For instance, for medStd, recall that, conditional on θ and δ, we can express the Standard SGP in the desired form $a = -(\theta + (1 - \lambda_1)\beta\delta)$ so $\sigma_a^2 = \nu_2 + \nu_3$, where $\nu_2 = (1 - \lambda_1)^2 \beta^2 \sigma_\delta^2$ and $\nu_3 = \sigma_\theta^2$, and, as we have already defined, $b^2 = \sigma_{W_2}^2 (1 - \rho^2)$, and $\sigma_\xi^2 = \nu_2$.

The second component of the unconditional MSE of a medGP estimator in Equation A.18 is the squared bias, $\left(E\left[E[\text{medGP} \mid \theta, \delta]^2\right] - E\left[E[\text{medGP} \mid \theta]^2\right]\right)$. This term can be found by applying the results for the expectation of the square of the normal CDF evaluated at a normal random variable given in Equation A.4. For instance, for the medStd, the $E\left[E[\text{medStd} \mid \theta, \delta]^2\right]$ is a bivariate normal probability $Pr[t_1 \leq 0, t_2 \leq 0]$, where t_1, t_2 are bivariate normal with mean vector $(0, 0)$ (as when not conditioning on any terms, there are no longer any fixed terms a in the numerator), variances $\sigma_{W_2}^2(1 - \rho^2) + \nu_2 + \nu_3$ and covariance $\nu_2 + \nu_3$, where

$\nu_2 = (1-\lambda_1)^2\beta^2\sigma_\delta^2$ and $\nu_3 = \sigma_\theta^2$. And similarly, the $E\left[E[\text{medStd} \mid \theta]^2\right]$ is a bivariate normal probability $Pr[t_1 \leq 0, t_2 \leq 0]$, where t_1, t_2 are bivariate normal with mean vector $(0,0)$, variances $\sigma_{W_2}^2(1-\rho^2) + \nu_2 + \nu_3$ and covariance ν_3, where $\nu_2 = (1-\lambda_1)^2\beta^2\sigma_\delta^2$ and $\nu_3 = \sigma_\theta^2$. Note that these results differs from those for the meanStd (given above in Section A.1.5), in that ν_1 does not appear in the variances or covariances.

A.2.6 var $(E[\text{medGP} \mid \theta])$

The variance of the target of each medGP estimator is $\text{var}(E[\text{medGP} \mid \theta])$. As shown in Section A.1.6, this quantity can be decomposed into two expectations:

$$Var[E(\text{medGP} \mid \theta)] \approx E\left[E(\text{medGP} \mid \theta)^2\right] - E\left[E(\text{medGP} \mid \theta)\right]^2, \tag{A.20}$$

which again follow from our two general results for normal distributions. For instance, for the medStd, when not conditioning on any terms, $E(\text{medStd} \mid \theta)$ can be written in our desired form, $\Phi\left(\frac{\xi-a}{b}\right)$, with $\xi = \theta$, $\sigma_\xi^2 = \sigma_\theta^2$, $a = 0$ (i.e., there are no fixed terms), and $b = \sqrt{\sigma_{W_2}^2(1-\rho^2) + \nu_2}$. The first term in the above equation follows from Equation A.4 and is derived in the previous Section A.2.5. The second term involves the expectation of the normal CDF evaluated at a normal random variable which follows from applying Equation A.1: $E\left[E(\text{medGP} \mid \theta)\right] = \Phi\left(\frac{0}{\sqrt{\sigma_{W_2}^2(1-\rho^2)+\nu_2+\nu_3}}\right) = \Phi(0) = .5$. That is, the unconditional expected value of the target is a medGP of 50.

A.2.7 Median Standard SGP

The first research question in the paper focused on determining the effect of ME on the bias and variance of Standard MGP. In the paper, we focus on the mean Standard, but here, we provide the corresponding results for the median Standard SGP. Figure A1 of variance and bias terms for the medStd and medTrue is comparable to Figure 2 for the meanStd and meanTrue. We use the same parameter values to generate this figure ($\lambda_1 = \lambda_2 = .9, n = 50, \rho = .775, \pi_\theta = .04, \pi_\delta = .35, \sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$). The general story here is the same as for the meanStd:

1. Panel (a) shows that the $\text{var}(E[\text{medStd} \mid \theta, \delta] \mid \theta)$ is entirely due to ME in prior test scores as this variance for medTrue is 0 for all values of θ.

2. Panel (b) shows that the $E\left[\text{var}\left(\text{medStd} \mid \theta, \delta\right) \mid \theta\right]$ is due primarily to sampling variance as this

expected variance is similar to that for medTrue, but it is inflated by test score ME as the black curve for medStd is higher than the grey curve for medTrue for all values of θ.

3. Panel (c) shows the combined effect of ME and sampling error on the var(medStd$\,|\,\theta$), which is the sum of the terms shown in panels (a) and (b).

4. Panel (d) shows the systematic bias in medStd due to prior score ME with teachers of high achieving classes advantaged and those of low achieving classes disadvantaged.

Although these general findings are the same, we see that these variance terms for medStd and medTrue are larger than those for meanStd and meanTrue given in Figure 2. For instance, the $E\left[\text{var}\left(\text{medGP} \mid \theta, \delta\right) \mid \theta\right]$ for the medTrue reaches a maximum of about 41.40 (at the 50th percentile of θ), whereas this expected variance is just a little more than a third of this size for the meanTrue at 14.96. This result is not surprising as medians are known to have larger sampling variability than means.

Because the sampling variability is so much larger for the medGP estimators, we find that the proportion of error variance due to ME (propME) is actually smaller for the medStd than it is for the meanStd as shown by comparing the curve for Standard MGP in panel (a) in Figure A2 to panel (a) in Figure 3. However, the proportion of overall variance (instead of just error variance) due to ME is higher for medStd than for meanStd. For instance, for $\lambda = .9$, about 6.2% of the overall variance in meanStd is due to ME compared to 7.3% for medStd. This occurs because although sampling variability is higher for medStd, the signal variance, or var(E[MGP$\,|\,\theta$]), is lower. The ratio of error variance to signal variance is thus higher for medStd than meanStd, resulting in lower PRMSEs for the median over the mean as seen by comparing panels (c) and (d) in Figure A2 to those in Figure 3.

Figure A2 also helps address research question 2 about the extent that alternatives improve upon the medStd. Just as with the comparison of the mean estimators, Ranked SIMEX outperforms the Standard. The medSIMEX also has higher PRMSE but its propME is higher than that for the medStd for high test reliabilities and low observed current score ICCs. Although again, if we considered the proportion of overall variance due to ME, the medSIMEX would outperform the medStd.

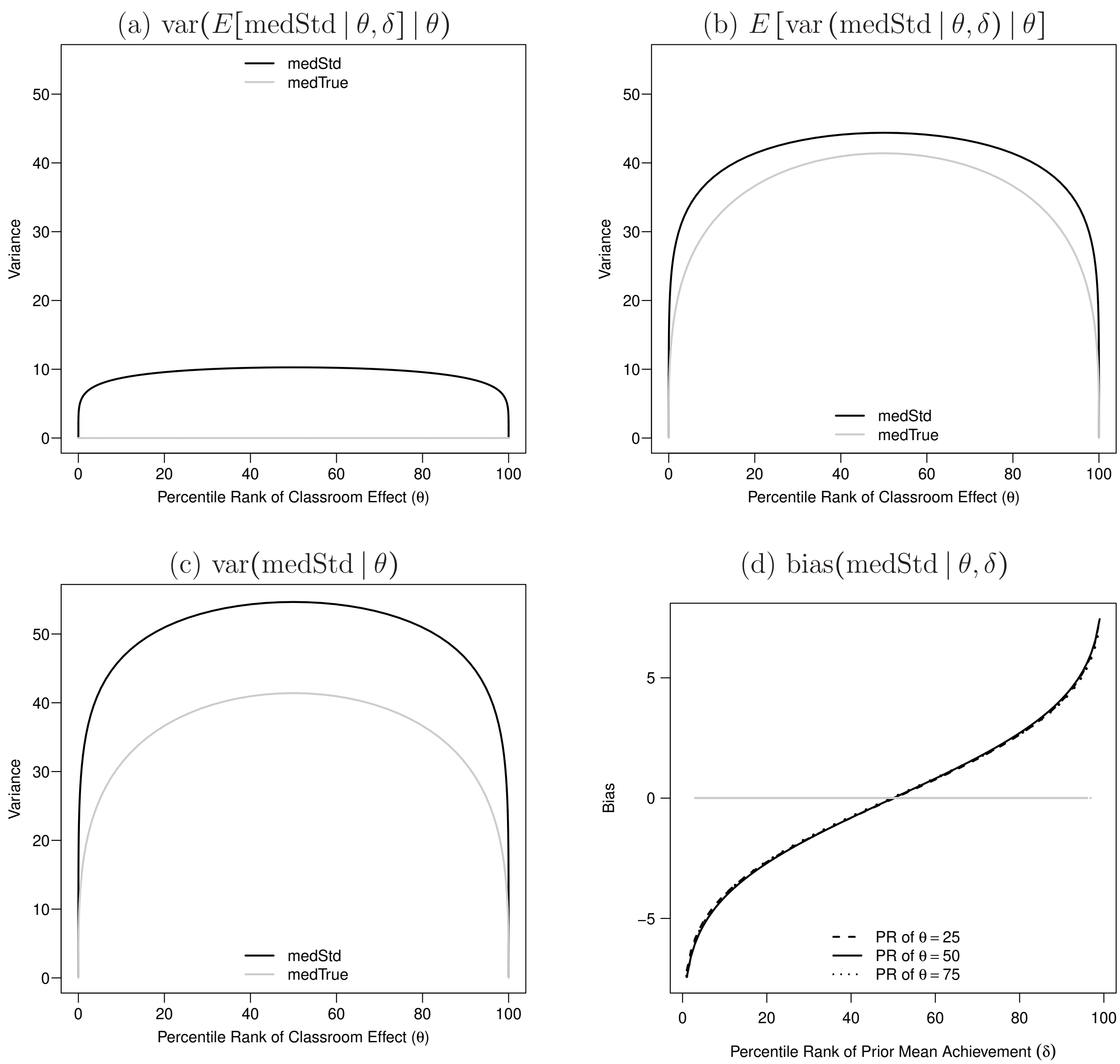

Figure A1: The effect of measurement error on the variance and bias in the median Standard SGP.

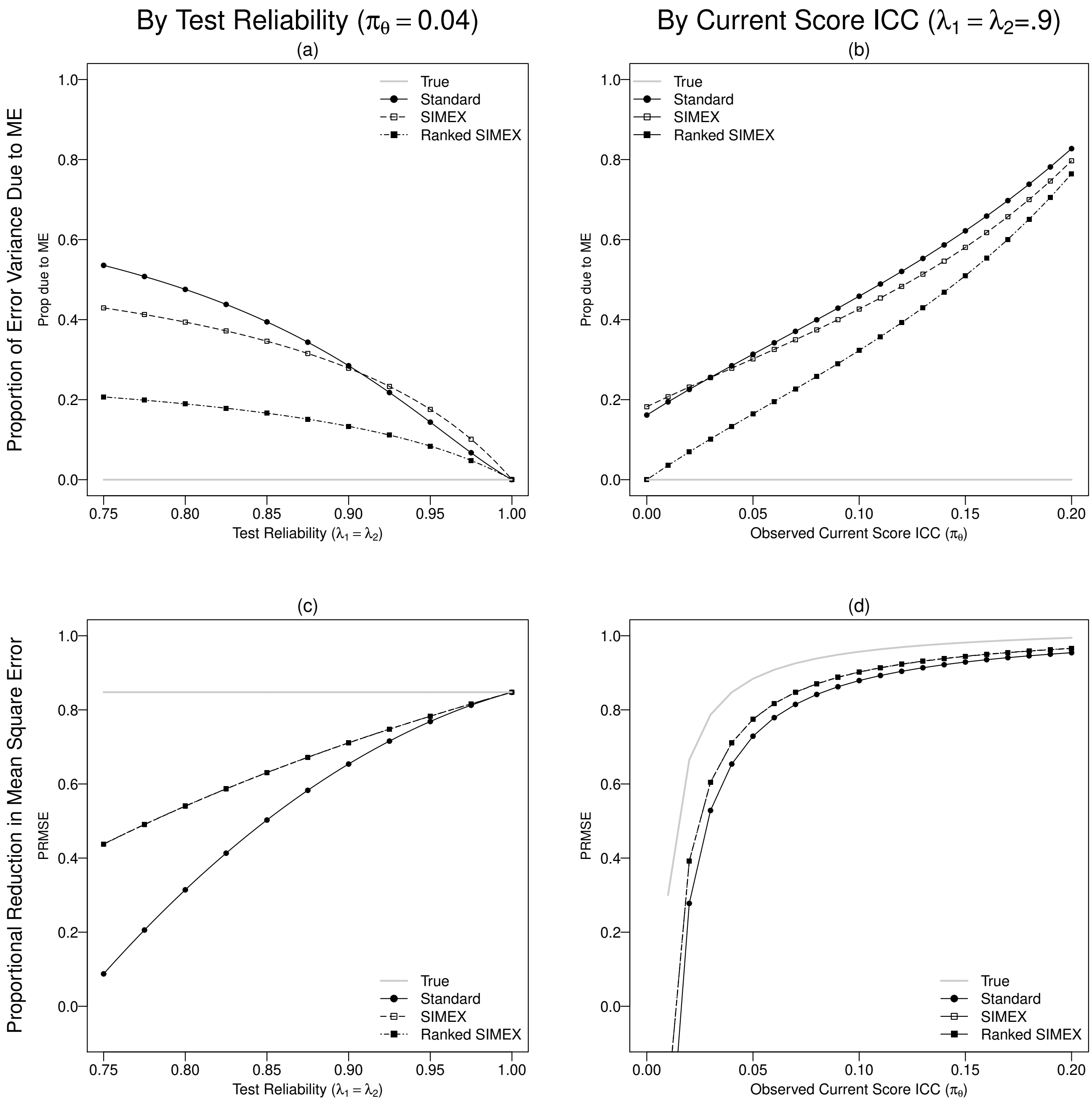

Figure A2: The extent that the effect of ME on the variance of medGP is modulated by varying test reliability ($\lambda_1 = \lambda_2$) and the observed score ICC (π_θ).

A.3 Nonzero Correlation between θ and δ

In the paper, we assume that $\mathrm{Cov}(\theta, \delta) = \omega = 0$ or that there is no teacher sorting with more/less effective teachers being assigned to more/less able students. However, this may not be the case in practice. If there is systematic sorting of teachers by their effectiveness level to classes by their prior mean ability level ($\omega \neq 0$), then the population regression coefficient from regressing true current score X_2 on true prior score X_1 becomes $\beta^* = \beta + \frac{\omega}{\sigma^2_{X_1}}$—it is no longer just β. Accordingly, ω will now play a part in our SGP formulas. Here we show their effect on the expected values of meanTrue and meanStd, which again follow from applying the general result for normal distributions given in Equation A.1.

For meanTrue, we have:

$$\begin{aligned}
E[\mathrm{meanTrue} \mid \theta, \delta] &= E[\mathrm{SGP_{True}} \mid \theta, \delta] \\
&= E\left[\Phi\left(\frac{X_2 - \beta^* X_1}{\sqrt{\sigma^2_{X_2}(1-\beta^{*2})}}\right)\middle| \theta, \delta\right] \\
&= E\left[\Phi\left(\frac{\theta + \eta - \left(\frac{\omega}{\sigma^2_{X_1}}\right)(\delta + \zeta)}{\sqrt{\sigma^2_{X_2}(1-\beta^{*2})}}\right)\middle| \theta, \delta\right] \\
&= \Phi\left(\frac{\theta - \left(\frac{\omega}{\sigma^2_{X_1}}\right)\delta}{\sqrt{\sigma^2_{X_2}(1-\beta^{*2}) + \sigma^2_\eta + (\frac{\omega}{\sigma^2_{X_1}})^2\sigma^2_\zeta}}\right).
\end{aligned} \tag{A.21}$$

and because $\delta = \frac{\omega}{\sigma^2_{X_1}}\theta + e$:

$$\begin{aligned}
E[\mathrm{meanTrue} \mid \theta] &= E\left[E[\mathrm{meanTrue} \mid \theta, \delta] \mid \theta\right] \\
&= E\left[\Phi\left(\frac{\theta - \left(\frac{\omega}{\sigma^2_{X_1}}\right)\delta}{\sqrt{\sigma^2_{X_2}(1-\beta^{*2}) + \sigma^2_\eta + \left(\frac{\omega}{\sigma^2_{X_1}}\right)^2\sigma^2_\zeta}}\right)\middle| \theta\right] \\
&= E\left[\Phi\left(\frac{\theta - \left(\frac{\omega}{\sigma^2_{X_1}}\right)\left(\frac{\omega}{\sigma^2_{X_1}}\theta + e\right)}{\sqrt{\sigma^2_{X_2}(1-\beta^{*2}) + \sigma^2_\eta + (\frac{\omega}{\sigma^2_{X_1}})^2\sigma^2_\zeta}}\right)\middle| \theta\right] \\
&= \Phi\left(\frac{\left(1 - \left(\frac{\omega}{\sigma^2_{X_1}}\right)^2\right)\theta}{\sqrt{\sigma^2_{X_2}(1-\beta^{*2}) + \sigma^2_\eta + \left(\frac{\omega}{\sigma^2_{X_1}}\right)^2\sigma^2_\zeta + \left(\frac{\omega}{\sigma^2_{X_1}}\right)^2\mathrm{var}(e)}}\right),
\end{aligned} \tag{A.22}$$

where $\text{var}(e) = \sigma_\delta^2(1-\varrho^2)$ and $\varrho = \text{Cor}(\theta,\delta) = \frac{\omega}{\sqrt{\sigma_\theta^2\sigma_\delta^2}}$. Accordingly, when there is teacher sorting, it is no longer the case that $E[\text{meanTrue} \mid \theta, \delta] = E[\text{meanTrue} \mid \theta]$. Thus, even when there is no ME, MGP will be biased in that the expected value of the meanTrue will be a function of δ, the students' mean prior achievement, and teachers will be advantaged or disadvantaged by a pre-existing characteristic of the students who they teach.

For meanStd, we obtain:

$$\begin{aligned}
E[\text{meanStd} \mid \theta, \delta] &= E[\text{SGP}_{\text{Std}} \mid \theta, \delta] \\
&= E\left[\Phi\left(\frac{W_2 - \lambda_1\beta^* W_1}{\sqrt{\sigma_{W_2}^2(1-\rho^2)}}\right)\middle| \theta, \delta\right] \\
&= E\left[\Phi\left(\frac{\theta + \eta + \left[(1-\lambda_1)\beta - \lambda_1\frac{\omega}{\sigma_{X_1}^2}\right](\delta+\zeta) - \lambda_1(\beta - \frac{\omega}{\sigma_{X_1}^2})U_1 + U_2}{\sqrt{\sigma_{W_2}^2(1-\rho^2)}}\right)\middle| \theta, \delta\right] \\
&= \Phi\left(\frac{\left[(1-\lambda_1)\beta - \lambda_1\frac{\omega}{\sigma_{X_1}^2}\right]\delta + \theta}{\sqrt{\sigma_{W_2}^2(1-\rho^2) + \tau_1}}\right), \qquad \text{(A.23)}
\end{aligned}$$

where $\tau_1 = \sigma_\eta^2 + \sigma_{U_2}^2 + \left[(1-\lambda_1)\beta - \lambda_1\frac{\omega}{\sigma_{X_1}^2}\right]^2\sigma_\zeta^2 + \lambda_1^2(\beta - \frac{\omega}{\sigma_{X_1}^2})^2\sigma_{U_1}^2$, and

$$\begin{aligned}
E[\text{meanStd} \mid \theta] &= E\left[E[\text{meanStd} \mid \theta, \delta] \mid \theta\right] \\
&= E\left[\Phi\left(\frac{\left[(1-\lambda_1)\beta - \lambda_1\frac{\omega}{\sigma_{X_1}^2}\right]\delta + \theta}{\sqrt{\sigma_{W_2}^2(1-\rho^2) + \tau_1}}\right)\middle| \theta\right] \\
&= E\left[\Phi\left(\frac{\left[(1-\lambda_1)\beta - \lambda_1\frac{\omega}{\sigma_{X_1}^2}\right]\left(\frac{\omega}{\sigma_{X_1}^2}\theta + e\right) + \theta}{\sqrt{\sigma_{W_2}^2(1-\rho^2) + \tau_1}}\right)\middle| \theta\right] \\
&= \Phi\left(\frac{\left[((1-\lambda_1)\beta - \lambda_1\frac{\omega}{\sigma_{X_1}^2})\frac{\omega}{\sigma_{X_1}^2} + 1\right]\theta}{\sqrt{\sigma_{W_2}^2(1-\rho^2) + \tau_1 + \tau_2}}\right), \qquad \text{(A.24)}
\end{aligned}$$

where $\tau_2 = ((1-\lambda_1)\beta - \lambda_1\frac{\omega}{\sigma_{X_1}^2})^2\text{var}(e)$. Note that if we plug in $\omega = 0$ for no teacher sorting, all of these formulas reduce to those provided in the paper.

B Results for Alternative MGPs

We described two alternatives to the Standard MGP estimators in the paper: SIMEX and Ranked SIMEX. Here, we expand on our description of the implementation of the SIMEX measurement error correction for SGP, and describe other possible alternatives in more detail. We then evaluate them in comparison to the Standard MGP. Table B1 summarizes all of the MGP estimators, classifying the alternatives into three classes: ME-corrected measures, measures using the true score distribution of (X_1, X_2), and direct measures, or measures that directly estimate True MGP, instead of estimating a student-level SGP first and then aggregating to the teacher (or school) level. Table B2 extends Table 1 given in the paper to include the other estimators. It also includes the asymptotic expectations for the corresponding median estimators. Note that these results follow using the formulas given in Sections A.1.1 and A.1.1 for the aggregated mean estimators and Sections A.2.1 and A.2.2 for the aggregated median estimators. In Section B.4, we note how these conditional expectations are obtained for the Direct measures. We re-arrange terms in the numerators and denominators of each expected value to facilitate direct comparison among the estimators, including what we would obtain if tests had no ME (i.e., the expectations for meanTrue and medTrue).

Table B1: *Descriptions of MGP Estimators.*

Class	Name	Brief Description	Student-level Dist Shape	Prior Score ME bias in MGP?
		Aggregated Student-Level Estimators		
Standard SGP	(1) Standard	Uses observed prior and current scores	Uniform	Yes
ME-corrected SGP	(2) SIMEX	Uses SIMEX to correct for ME in prior scores	U-shaped	No
	(3) Ranked SIMEX	Percentile rank of SIMEX	Uniform	No
	(4) Double-Corrected	Uses SIMEX to correct for ME in prior and current scores	U-shaped	No
SGP using the Dist of (X_1, X_2)	(5) MLE	Uses the MLEs of X_1 and X_2 in the true distribution of X_1 and X_2.	U-shaped	No
	(6) EAP	Equals the posterior mean $E[\text{SGP}_{\text{True}} \mid W_1, W_2]$	Bell-shaped	Yes
		Direct MGP Estimators		
Direct Measures	(7) DirectMLE	Uses the MLE of θ to directly estimate the True MGP.	N/A	Yes
	(8) DirectEAP	Equals the posterior mean $E[\text{MGP}_{\text{True}} \mid W_1, W_2]$ to directly estimate the True MGP	N/A	Yes

Table B2: *MGP Estimators*

		Mean Estimators		Median Estimators	
Estimator	Student-Level	$E[\text{meanEst} \mid \theta, \delta]$	$E[\text{meanEst} \mid \theta]$	$E[\text{medEst} \mid \theta, \delta]$	$E[\text{medEst} \mid \theta]$
(0) True	$\Phi\left(\frac{\epsilon}{\sqrt{\sigma_\epsilon^2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2}}\right)$	$= E[\text{meanTrue} \mid \theta, \delta]$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2}}\right)$	$= E[\text{medTrue} \mid \theta, \delta]$
(1) Standard	$\Phi\left(\frac{\epsilon+U_2-\lambda_1\beta U_1+(1-\lambda_1)\beta X_1}{\sqrt{\sigma_\epsilon^2-\sigma_\eta^2+\nu_1+\nu_2}}\right)$	$\Phi\left(\frac{\theta+(1-\lambda_1)\beta\delta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+2(\nu_1-\sigma_\eta^2)+\nu_2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+2(\nu_1-\sigma_\eta^2)+2\nu_2}}\right)$	$\Phi\left(\frac{\theta+(1-\lambda_1)\beta\delta}{\sqrt{\sigma_\epsilon^2-\sigma_\eta^2+\nu_1+\nu_2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2-\sigma_\eta^2+\nu_1+2\nu_2}}\right)$
(2) SIMEX	$\Phi\left(\frac{\epsilon+U_2-\beta U_1}{\sqrt{\sigma_\epsilon^2+\sigma_{U_2}^2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+2\sigma_{U_2}^2+\beta^2\sigma_{U_1}^2}}\right)$	$= E[\text{meanSIMEX} \mid \theta, \delta]$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_{U_2}^2}}\right)$	$= E[\text{medSIMEX} \mid \theta, \delta]$
(3) Ranked SIMEX (RS)	$\Phi\left(\frac{\epsilon+U_2-\beta U_1}{\sqrt{\sigma_\epsilon^2+\sigma_{U_2}^2+\beta^2\sigma_{U_1}^2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+2\sigma_{U_2}^2+2\beta^2\sigma_{U_1}^2}}\right)$	$= E[\text{meanRS} \mid \theta, \delta]$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_{U_2}^2+\beta^2\sigma_{U_1}^2}}\right)$	$= E[\text{medRS} \mid \theta, \delta]$
(4) Double-Corrected (DC)	$\Phi\left(\frac{\epsilon+U_2-\beta U_1}{\sqrt{\sigma_\epsilon^2-\sigma_{U_2}^2-\beta^2\sigma_{U_1}^2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2}}\right)$	$= E[\text{meanDC} \mid \theta, \delta]$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2-\sigma_{U_2}^2-\beta^2\sigma_{U_1}^2}}\right)$	$= E[\text{medDC} \mid \theta, \delta]$
(5) MLE	$\Phi\left(\frac{\epsilon+U_2-\beta U_1}{\sqrt{\sigma_\epsilon^2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+\sigma_{U_2}^2+\beta^2\sigma_{U_1}^2}}\right)$	$= E[\text{meanMLE} \mid \theta, \delta]$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2}}\right)$	$= E[\text{medMLE} \mid \theta, \delta]$
(6) EAP	$\Phi\left(\frac{\psi[\epsilon+U_2+\beta U_1+(1-\lambda_1)\beta X_1]}{\sqrt{\sigma_\epsilon^2(2-\psi)}}\right)$	$\Phi\left(\frac{\psi[\theta+(1-\lambda_1)\beta\delta]}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+\sigma_\theta^2-\psi\sigma_\epsilon^2+\psi^2\nu_1}}\right)$	$\Phi\left(\frac{\psi\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+\sigma_\theta^2-\psi\sigma_\epsilon^2+\psi^2\nu_2}}\right)$	$\Phi\left(\frac{\psi[\theta+(1-\lambda_1)\beta\delta]}{\sqrt{\sigma_\epsilon^2(2-\psi)}}\right)$	$\Phi\left(\frac{\psi\theta}{\sqrt{\sigma_\epsilon^2(2-\psi)+\psi^2\nu_2}}\right)$
(7) DirectMLE	N/A	$\Phi\left(\frac{\theta+(1-\lambda_1)\beta\delta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+\sigma_\kappa^2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+\sigma_\kappa^2+\nu_2}}\right)$	$\Phi\left(\frac{\theta+(1-\lambda_1)\beta\delta}{\sqrt{\sigma_\epsilon^2+\sigma_\kappa^2}}\right)$	$\Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\kappa^2+\nu_2}}\right)$
(8) DirectEAP	N/A	$\Phi\left(\frac{\Lambda[\theta+(1-\lambda_1)\beta\delta]}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+\sigma_\theta^2(1-\Lambda)+\Lambda^2\sigma_\kappa^2}}\right)$	$\Phi\left(\frac{\Lambda\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2+\sigma_\theta^2(1-\Lambda)+\Lambda^2\sigma_\kappa^2+\Lambda^2\nu_2}}\right)$	$\Phi\left(\frac{\Lambda[\theta+(1-\lambda_1)\beta\delta]}{\sqrt{\sigma_\epsilon^2+\sigma_\theta^2(1-\Lambda)+\Lambda^2\sigma_\kappa^2}}\right)$	$\Phi\left(\frac{\Lambda\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\theta^2(1-\Lambda)+\Lambda^2\sigma_\kappa^2+\Lambda^2\nu_2}}\right)$

Note. $\nu_1 = \sigma_\eta^2 + \sigma_{U_2}^2 + (1-\lambda_1)^2\beta^2\sigma_\zeta^2 + \lambda_1^2\beta^2\sigma_{U_1}^2$; $\nu_2 = (1-\lambda_1)^2\beta^2\sigma_\delta^2$; $\psi = \sigma_\epsilon^2/(\sigma_{W_1}^2(1-\rho^2))$; $\Lambda = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\kappa^2+(1-\lambda_1)^2\beta^2\sigma_\delta^2}$

B.1 ME-Corrected Measures

In this section we provide details on three alternative MGP estimators that rely on SIMEX to correct the standard SGP methods for measurement error in the tests. These include the SIMEX and Ranked SIMEX MGP estimators discussed in the paper and a third MGP estimtator that falls into this class–the Double-Corrected MGP.

B.1.1 SIMEX

Simulation-Extrapolation (SIMEX) is a ME correction technique which is particularly useful for nonlinear models such as the quantile models commonly used in estimating the Standard SGP. SIMEX involves two steps: 1) *Simulation*: Extra measurement error with variance $(1+\gamma)\sigma^2_{U_1}$ is added to the observed data repeatedly B times and the model is estimated using standard methods that ignore measurement error. This step is repeated at multiple values of $\gamma > 0$. 2) *Extrapolation*: At each value of γ, the estimates are averaged over B replications, and the average estimates are modeled as function of γ. The model is then projected back to $\gamma = -1$, or the value where there is no measurement error. The projected value is the final SIMEX estimate. If measurement error is affecting the estimates, then increasing the measurement variance (by a factor of $1+\gamma$) should amplify the effects and the estimates should change as a function of γ. At $\gamma = -1$, the measurement error variance would be zero, so there would be no measurement error, and this value should be a consistent estimate of the parameters of interest. SIMEX requires estimating the projection function correctly and that the parameter estimates be a smooth monotonic function of γ. The accuracy of the SIMEX estimate depends on the number of γ values, the number or replications B, the sample sized use to estimate the parameters of interest, and the accuracy of the projection function. We assume the sample size and B are essentially infinite and that the projection model is known exactly. Of course, these conditions will not hold in practice so our results are a best case for SIMEX estimators.

Shang et al. (2015) use SIMEX to estimate the quantile regression functions, which they then use to create the conditional distribution of W_2. Using SIMEX to project the quantiles by adding noise with variance $(1+\gamma)\sigma^2_{U_1}$ to each W_1, repeatedly estimating the quantile functions at the observed scores (with extra noise added) B times, and taking the average across the B replicates for a given student, yields an estimated function for the p^{th} quantile, $p \in (0,1)$, approximately equal to:

$$p_\gamma(W_1) = z_p\sqrt{\sigma_\epsilon^2 + (1-\lambda_2)\sigma_{W_2}^2 + (1-\lambda_\gamma)\lambda_\gamma\beta^2\sigma_{W_1}^2} + \lambda_\gamma\beta E[W_1^* \mid W_1], \tag{B.1}$$

Where, z_p is the p^{th} quantile of the standard normal distribution, W_1^* is the simulated score with extra noise, $\lambda_\gamma = \sigma_{X_1}^2/(\sigma_{X_1}^2 + (1+\gamma)\sigma_{U_1}^2)$ which converges to 1 as γ goes to -1, and $E[W_1^* \mid W_1] = W_1$. Thus, as γ converges to -1, $p_\gamma(W_1)$ converges to $z_p\sqrt{\sigma_\epsilon^2 + (1-\lambda_2)\sigma_{W_2}^2} + \beta W_1$, the p^{th} quantile of normal distribution with mean βW_1 and variance $\sigma_\epsilon^2 + (1-\lambda_2)\sigma_{W_2}^2$. Conditional on X_1, $W_2 \sim N(\beta X_1, \sigma_\epsilon^2 + (1-\lambda_2)\sigma_{W_2}^2)$, so SIMEX yields an estimate of $F_{W_2|X_1}(W_2 = w_2 \mid X_1 = w_1)$. Note, that variance of $W_2 - \beta W_1$ is $\sigma_\epsilon^2 + (1-\lambda_2)\sigma_{W_2}^2 + \beta^2\sigma_{U_1}^2 > \sigma_\epsilon^2 + (1-\lambda_2)\sigma_{W_2}^2$, so that distribution of the SIMEX SGP is U-shaped rather than uniform.

B.2 Double-Corrected

A third possible SIMEX estimator results from correcting for ME in both the prior and current observed scores. McCaffrey et al. (2015) propose such an estimator, which we refer to as the "Double Corrected" SGP (DC).[7] Under our distributional assumptions, this SGP estimator has the same numerator term as SIMEX SGP, $W_2 - \beta W_1$, but involves subtracting variance terms in the denominator instead of summing them:[8]

$$\mathrm{SGP}_{\mathrm{DC}} = \Phi\left(\frac{W_2 - \beta W_1}{\sqrt{\sigma_\epsilon^2 - \sigma_{U_2}^2 - \beta^2\sigma_{U_1}^2}}\right). \tag{B.2}$$

This subtraction in the denominator is a result of correcting for the ME in the current score and results in the Double-Corrected SGP estimator having an extremely U–shaped distribution with students essentially all being assigned an SGP of 0 or 100 with equal probability. Note also that the Double-Corrected SGP may be particularly difficult to estimate in practice given it involves applying SIMEX to adjust for ME in both the prior and current scores.

The expected mean and median Double-Corrected SGP given true classroom effects and prior class achievement have some interesting features. As shown in Table B2, the expected value for the meanDC is

[7] McCaffrey et al. (2015) refer to this estimator as the "Unbiased SGP" because at the student-level, it is an unbiased estimator of True SGP given true prior and current achievement. However, we use $E[\mathrm{MGP} \mid \theta]$ as opposed to meanTrue or medTrue as the target and thus the label unbiased could be confusing. Accordingly, we use the more descriptive term "Double Corrected."

[8] Due to the subtraction in the expression within the square root of this SGP estimator, $\mathrm{SGP}_{\mathrm{DC}}$ only exists under certain conditions; specifically, if $\rho^2 < \lambda_1^2(2\lambda_2 - 1)$ and $\lambda_2 > .5$.

equivalent to that of the meanTrue so that the expectation of the meanDC for teachers will equal the same as what we would obtain if the test scores had no ME. This occurs because, conditional on θ and δ, SGP_{DC} can be expressed in the desired form for our general result for the expectation of $\Phi\left(\frac{\xi-a}{b}\right)$, with $a = -\theta$ (the fixed component), $\xi = \eta + U_2 + \beta U_1$ (the random component) with variance $\sigma_\xi^2 = \sigma_\eta^2 + \sigma_{U_2}^2 + \beta^2\sigma_{U_1}^2$, and $b^2 = \sigma_\epsilon^2 - \sigma_{U_2}^2 - \beta^2\sigma_{U_1}^2$, so that $E\left[\Phi\left(\frac{\xi-a}{b}\right)\right] = \Phi\left(\frac{-a}{\sqrt{b^2+\sigma_\xi^2}}\right) = \Phi\left(\frac{\theta}{\sqrt{[\sigma_\epsilon^2-\sigma_{U_2}^2-\beta^2\sigma_{U_1}^2]+[\sigma_\eta^2+\sigma_{U_2}^2+\beta^2\sigma_{U_1}^2]}}\right) = \Phi\left(\frac{\theta}{\sqrt{\sigma_\epsilon^2+\sigma_\eta^2}}\right)$. In contrast, the asympototic expected value of medDC does not equal that of the medTrue as seen by comparing the asymptotic conditional expectation of medDC to medTrue in Table B2. Moreover, $E[\text{medDC} \mid \theta, \delta]$ includes subtraction in its denominator like the student-level Double-Corrected SGP, which will prove to be problematic and yield undesirable properties compared to the other alternative median estimators.

B.3 SGP Using the Distribution of (X_1, X_2)

The standard approach to estimating SGP and the SIMEX corrected versions, typically make no explicit assumptions about the joint distribution for the true scores, X_1 and X_2. If assumptions about that distribution are used, then additional SGP estimators become available. One such approach entails estimating $F_{X_2|X_1}$ using the assumed functional form. Lockwood and Castellano (2015) and Monroe and Cai (2015) suggest using longitudinal item-level data and multidimensional item response theory models (MIRT) to obtain more accurate estimates of the CDF of interest ($F_{X_2|X_1}$), rather than using ME corrections and the observed test scores. Lockwood and Castellano (2015) discuss possible deconvolution methods to estimate the conditional distributions of the true scores from the observed item response data. Alternative deconvolution methods can be used to obtain an estimate of the joint distribution using just the observed test scores and information about the ME in those scores (Carroll et al., 2006). Once an estimate of the CDF $F_{X_2|X_1}$ has been obtained, it can then be used to estimate the SGP by evaluating it at estimates for X_1 and X_2.

An obvious choice for estimates of X_1 and X_2 are W_1 and W_2. Provided the test scores are MLE estimates of X_1 and X_2, then $F_{W_2|W_1}$ is the MLE of the true SGP. Hence we call this estimator the MLE

SGP (SGP_{MLE}). Under our distributional assumption, this yields:

$$\begin{aligned} \text{SGP}_{\text{MLE}} &= F_{X_2|X_1}(X_2 = w_2 \mid X_1 = w_1) \\ &= \Phi\left(\frac{W_2 - \beta W_1}{\sqrt{\sigma_\epsilon^2}}\right). \end{aligned} \tag{B.3}$$

Like the SIMEX and Double-Corrected SGP, the MLE SGP has a U-shaped distribution (though not as extreme as the Double–Corrected's U-shaped distribution). Inspection of Equation B.3 reveals that the numerator of MLE SGP equals the numerator in the Ranked SIMEX SGP. Ranking an SGP estimator results in a uniformly distribute estimator, which means that for a ranked SGP estimator, the denominator of the quantity in the normal CDF must equal the standard deviation of the numerator. Since, the numerator of the Ranked SIMEX equals the numerator of MLE SGP, the denominator of the Ranked SIMEX must equal the standard deviation of the numerator of MLE SGP, and a Ranked MLE SGP would equal the Ranked SIMEX SGP.

The numerator of MLE SGP is the same as that for all the estimators in the ME-corrected class (SIMEX, Ranked SIMEX and Double Corrected). Accordingly, the numerators of the expected mean and median MLE SGP (given true teacher effects and prior class achievement) are also the same as those for the ME-corrected MGP as shown in Table B2. That is, all of these MGP estimators, like the True MGP, are nonlinear, monotonic functions of the true classroom effect and do not depend on prior class achievement. As seen in Table B2, the asymptotic expected medMLE equals that for medTrue. Thus, for large class sizes, the median MLE SGP will be, on average, equal to what we would obtain if there was no ME in the prior or current test scores. This is not the case for the meanMLE.

Alternatively, rather than evaluating the CDF at the observed test scores, we can use it as the prior for the unknown true scores and estimate the EAP SGP, as the posterior mean of the True SGP given the observed data as in McCaffrey et al. (2015):

$$
\begin{aligned}
\text{SGP}_{\text{EAP}} &= E[\text{SGP}_{\text{True}} \mid W_1, W_2] \\
&= E\left[\Phi\left(\frac{(\epsilon-\mu)+\mu}{\sigma_\epsilon}\right) \mid W_1, W_2\right] \\
&= \Phi\left(\frac{\mu}{\sqrt{\sigma_\epsilon^2+\sigma_\epsilon^2(1-\psi)}}\right) \\
&= \Phi\left(\frac{\psi[\theta+\eta+U_2+\beta(1-\lambda_1)(\delta+\zeta)-\beta\lambda_1 U_1]}{\sqrt{(\sigma_\theta^2+\sigma_\eta^2)(2-\psi)}}\right),
\end{aligned}
\tag{B.4}
$$

where $\epsilon \mid W_1, W_2 \sim N(\mu, \sigma^2(1-\psi))$, $\mu = \psi W_2 - \frac{\sigma_{W_2}}{\sigma_{W_1}}\psi W_1$, and $\psi = \sigma_\epsilon^2/(\sigma_{W_1}^2(1-\rho^2))$. The last line of Equation B.4 follows by plugging in our decomposition of student-level error ϵ, $\epsilon = \theta + \eta$, and prior student achievement X_1, $X_1 = \delta + \zeta$, into their respective classroom and random student components. By definition, the EAP SGP minimizes mean square error, but it does so through shrinkage and is thus underdispersed relative to the distribution of True SGP. In other words, as indicated in Table B1, it has a bell-shaped distribution instead of the, arguably, more desirable uniform distribution for a student-level percentile rank measure. Because $\theta + \eta = \epsilon$, the value in the square bracket in the numerator of the quantity inside the normal CDF in the last line of Equation B.4 equals the numerator of the Standard SGP. Thus, ranking the EAP SGP will yield an estimator equal to the Standard SGP, since the Standard SGP is uniformly distributed and the Standard and EAP SGP share a common numerator, except for multiplication by a constant.

The expectations for meanEAP and medEAP given θ and δ, like those for meanStd and medStd, include both θ and δ in their numerators; thus, they are also correlated with mean prior achievement. Accordingly, the meanEAP and medEAP suffer from ME bias in the prior scores as indicated in Table B1. In addition, comparing these expected values to those for the Standard MGP, we see that the numerators are the same except that the EAP MGP's are multiplied by a shrinkage factor ψ.

B.4 Direct measures

The previous two estimators, MLE and EAP MGP, used assumptions about the joint distribution of X_1 and X_2 as a means of developing alternative SGP estimates that were then aggregated into MGP. Alternatively, we can try to obtain direct estimates of the expected values for meanTrue and medTrue given θ.

The observed test scores can be used to fit a hierarchical linear model for the current and prior scores. This model can be used to specify meanTrue and medTrue as functions of the classroom effect. Obtaining a direct estimate requires estimating the unknown θ. One possible direct estimator is the MLE which replaces the unknown θ with its MLE. Assuming no correction for ME, in a hierarchical linear model, the standard estimator of θ would be $\bar{W}_2 - \lambda_1\beta\bar{W}_1$. Hence the direct MLE estimators are:

$$\begin{aligned}\text{meanDirectMLE} &= \Phi\left(\frac{\bar{W}_2 - \lambda_1\beta\bar{W}_1}{\sqrt{\sigma_\theta^2 + 2\sigma_\eta^2}}\right) \\ &= \Phi\left(\frac{\theta + \bar{\eta} + (1-\lambda_1)\beta(\delta + \bar{\zeta}) + \bar{U}_2 - \lambda\beta\bar{U}_1}{\sqrt{\sigma_\theta^2 + 2\sigma_\eta^2}}\right), \text{ and} \qquad \text{(B.5)}\end{aligned}$$

$$\begin{aligned}\text{medDirectMLE} &= \Phi\left(\frac{\bar{W}_2 - \lambda_1\beta\bar{W}_1}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2}}\right) \\ &= \Phi\left(\frac{\theta + \bar{\eta} + (1-\lambda_1)\beta(\delta + \bar{\zeta}) + \bar{U}_2 - \lambda\beta\bar{U}_1}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2}}\right). \qquad \text{(B.6)}\end{aligned}$$

The direct estimator for the medTrue does not involve a sample median. Rather it involves averages of observed test scores for students in a teacher's class. If users feel that averaging SGP is inappropriate (due to the ordinal nature of percentile ranks), even for a population, then direct estimation provides a means of estimating median SGP potentially more efficiently using sample means of student test scores rather than medians of estimated SGP.

The expected value of the direct MLE estimators given θ and δ is straightforward to calculate under normality assumptions using our standard formulas for the expected value of the normal CDF evaluated at a normal random variable (Equation A.1; that is, even for the expected value of the median DirectMLE, we use the formula for the expected value of a meanGP given in Section A.1.1 as opposed to the result for the median aggregated SGPs given in Section A.2.1). Let $\kappa = \bar{\eta} + (1-\lambda_1)\beta\bar{\zeta} + \bar{U}_2 - \lambda\beta\bar{U}_1$ with variance

$\sigma_\kappa^2 = n^{-1}(\sigma_\eta^2 + (1-\lambda_1)^2\beta^2\sigma_\zeta^2 + \sigma_{U_2}^2 + \lambda_1^2\beta^2\sigma_{U_1}^2)$. Then

$$E\left[\text{meanDirectMLE} \mid \theta, \delta\right] = E\left[\Phi\left(\frac{\theta + (1-\lambda_1)\beta\delta + \kappa}{\sqrt{\sigma_\theta^2 + 2\sigma_\eta^2}}\right)\middle| \theta, \delta\right] = \Phi\left(\frac{\theta + (1-\lambda_1)\beta\delta}{\sqrt{\sigma_\theta^2 + 2\sigma_\eta^2 + \sigma_\kappa^2}}\right) and \tag{B.7}$$

$$E\left[\text{medDirectMLE} \mid \theta, \delta\right] = E\left[\Phi\left(\frac{\theta + (1-\lambda_1)\beta\delta + \kappa}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2}}\right)\middle| \theta, \delta\right] = \Phi\left(\frac{\theta + (1-\lambda_1)\beta\delta}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2 + \sigma_\kappa^2}}\right). \tag{B.8}$$

An alternative to plugging the MLE for θ into the formulas for meanTrue and medTrue is to estimate the EAP for the desired true MGP given the sample of test scores $\{W_1, W_2\}$ for a teacher. Treating the variance components and the estimated slope parameter as known (i.e., using empirical Bayes), we need to estimate the distribution of θ given $\{W_1, W_2\}$ and then find the expected value. Let $\Lambda = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\kappa^2 + (1-\lambda_1)^2\beta^2\sigma_\delta^2}$, then

$$\theta \mid \{W_1, W_2\} \sim N\left[\Lambda\left(\bar{W}_2 - \lambda_1\beta\bar{W}_1\right), \sigma_\theta^2(1-\Lambda)\right]. \tag{B.9}$$

The expected mean and median of θ given $\{W_1, W_2\}$ then follow using our standard formulas for expected values and $\sigma_\theta^2 + 2\sigma_\eta^2$ or $\sigma_\theta^2 + \sigma_\eta^2$ in the denominators for the mean and median respectively:

$$\text{meanDirectEAP} = \Phi\left(\frac{\Lambda\left(\bar{W}_2 - \lambda_1\beta\bar{W}_1\right)}{\sqrt{\sigma_\theta^2 + 2\sigma_\eta^2 + \sigma_\theta^2(1-\Lambda)}}\right) \tag{B.10}$$

$$\text{medDirectEAP} = \Phi\left(\frac{\Lambda\left(\bar{W}_2 - \lambda_1\beta\bar{W}_1\right)}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2 + \sigma_\theta^2(1-\Lambda)}}\right). \tag{B.11}$$

The expected value of the direct EAP estimators given θ and δ are straightforward to calculate given that $\bar{W}_2 - \lambda_1 \beta \bar{W}_1 = \theta + (1 - \lambda_1)\beta\delta + \kappa$. This yields

$$\begin{aligned} E\left[\text{meanDirectEAP} \mid \theta, \delta\right] &= E\left[\Phi\left(\frac{\Lambda\left(\theta + (1-\lambda_1)\beta\delta + \kappa\right)}{\sqrt{\sigma_\theta^2 + 2\sigma_\eta^2 + \sigma_\theta^2(1-\Lambda)}}\right)\middle| \theta, \delta\right] \\ &= \Phi\left(\frac{\Lambda\left(\theta + (1-\lambda_1)\beta\delta\right)}{\sqrt{\sigma_\theta^2 + 2\sigma_\eta^2 + \sigma_\theta^2(1-\Lambda) + \Lambda^2\sigma_\kappa^2}}\right). \end{aligned} \quad \text{(B.12)}$$

$$\begin{aligned} E\left[\text{medDirectEAP} \mid \theta, \delta\right] &= E\left[\Phi\left(\frac{\Lambda\left(\theta + (1-\lambda_1)\beta\delta + \kappa\right)}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2 + \sigma_\theta^2(1-\Lambda)}}\right)\middle| \theta, \delta\right] \\ &= \Phi\left(\frac{\Lambda\left(\theta + (1-\lambda_1)\beta\delta\right)}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2 + \sigma_\theta^2(1-\Lambda) + \Lambda^2\sigma_\kappa^2}}\right). \end{aligned} \quad \text{(B.13)}$$

Note that both of these direct estimators, the DirectMLE and DirectEAP MGP, depend on both θ and δ like the aggregated Standard and EAP SGP. Thus, although they are more direct in their estimation of the true MGP, they are correlated with prior mean achievement as they do not involve any corrections for ME in the current or prior scores. It is also useful to note the derivations for expected values and variances of the mean and median direct estimators are similar to those use to obtain the properties of the mean aggregated-student estimators given in Section A.1, but for the direct estimators, there are no corresponding student-level estimators. For the direct MGP, the references to student-level SGP estimators in Section A.1 can be replaced with the direct estimates themselves (Equations B.5, B.6, B.10, and B.11). In addition, when applying the formulas from Section A.1 to the direct MGP estimators, any instances of multiplying by $\frac{1}{n}$ can be ignored as they result from taking a sample mean of student SGP, which is not applicable for the direct measures. Instead, care must be taken to account for the fact that the direct estimators rely on the classroom averages of the current and prior year observed scores.

B.5 Evaluating the Alternatives

To evaluate the MGP estimators as alternatives to the Standard MGP, we compare the PRMSE given in Equation 9 in the paper as an overall summary of error in the estimators. The PRMSE quantifies the extent to which each MGP estimator reduces overall error variance compared to using the expected value of its target $E[\text{MGP} \mid \theta]$, which reduces to assigning all teachers an MGP of 50. Figure B1(a) reproduces the PRMSE curves as a function of test reliabilities ($\lambda_1 = \lambda_2$) given in Figure 3 for the True, Standard, SIMEX, and Ranked SIMEX meanGP and adds the corresponding PRMSE curves for the other alternative estimators. Panel (b) plots the PRMSE curves for the corresponding median estimators. The different types of MGP estimators are identified by different line types and plotting symbols. Although it is often difficult to distinguish among the curves, the fact that they often overlap indicates that the estimators or groups of estimators generally perform similarly by this overall outcome statistic.

There are a few noteworthy distinctions among the estimators' PRMSE. We first consider the aggregated-student MGP estimators (estimators 1 to 6). Comparing across panels, it is clear that the mean estimators have higher PRMSEs than their corresponding median estimators. For instance, for test reliabilities of .9, the PRMSE for meanStd is .75, while it is only .65 for the medStd. Thus, simply using the mean aggregation function instead of the median aggregation function can noticeably improve marginal reliability of the MGP esitmator.

Across these six estimators, the two alternatives considered in the paper, SIMEX (open square) and Ranked SIMEX (closed square), perform the best for both means and medians. However, MLE (star) performs almost comparably to SIMEX and Ranked SIMEX, particularly for the medians. It is not surprising that these are the three aggregated MGP estimators that have the highest PRMSE, or marginal reliability.They are independent of prior mean achievement δ so that they are unbiased by ME and do not disadvantage or advantage teachers based on the prior mean achievement of the teacher's class, and this bias is the primary means by which ME contributes to error in the Standard SGP.

The Double-Corrected ("+") is also independent of δ, but neither the meanDC nor medDC perform as well as the other MGP alternatives that remove the bias with prior mean achievement because of its unique properties. Recall that SGP_{DC} has an extremely U-shaped distribution due to subtraction of several variance terms in its denominator.[9] This distribution results in large variance for the Double-Corrected SGP and, subsequently, large sampling error in Double-Corrected MGP. The performance of the medDC is much worse than that of the meanDC. For instance, the meanDC's PRMSE at test reliabilities of .87 is .70, while that for the medDC at the same test reliability is only .51. Hence, at this moderately high

[9]Recall that due to the subtraction in the square root of the denominator of the Double-Corrected estimator, the SGP_{DC} does not exist if $\rho^2 > \lambda_1^2(2\lambda_2 - 1)$ or if $\lambda_2 < .5$ and thus it (and its mean and median) exists over a narrower range of test reliabilities than the other estimators in Figure B1.

test reliability, using the medDC only reduces MSE about 50% over using the trivial average estimator of assigning all teachers an MGP of 50.

The Standard (closed circle) and EAP (open circle) MGP estimators are the two other aggregated MGP estimators (as opposed to Direct measures). These two estimators perform moderately compared to the others though as test reliability increases, the distinction among them and the top performing estimators decreases.

As shown in Figure B1, the direct MGP estimators (estimators 7 and 8) perform differently among the means than the medians. The DirectMLE and DirectEAP mean estimators are nearly indistinguishable from the (aggregated-student) meanEAP (6) and are very similar to the meanStd (1). Thus, all four mean estimators that depend on δ have comparable overall error. In contrast, among the median estimators, the DirectMLE and DirectEAP appear to perform substantially better for test reliabilities of .85 or higher, even surpassing the medTrue at test reliabilities of .925 or higher. Comparing across panels (a) and (b), the curves for the DirectMLE and DirectEAP are nearly identical for means and medians. As noted in the previous section, this result follows from the direct "median" estimators relying on the sample means of test scores as opposed to using sample medians. Thus, they are more distinct from the median aggregated student SGP estimators than the direct "mean" estimators are from the mean aggregated student SGP estimators.

Figure B1 provides a direct comparison of each of the meanGP and medGP estimators. Although this figure only involves varying one of the parameters–test reliability–the relative order of the estimators will generally not change if we varied other parameters. The SIMEX and Ranked SIMEX would consistently outperform the others, followed closely by MLE MGP, while the Standard MGP lags behind. Given the parameter values used to generate this figure, at test reliabilities of .9, there are not substantial differences among the MGP estimators: the PRMSE of the mean aggregated student-level estimators and direct mean estimators are all within about .06 units of each other, ranging from .75 (the Standard and EAP) to .81 (Ranked SIMEX). Similarly, the median aggregated-student estimators are also within about .06 units, ranging from .65 (the Standard and EAP) to .71 (SIMEX and Ranked SIMEX), which increases to about .11 units if we include the direct "median" estimators that both have PRMSE of about .76. However, the SIMEX, Ranked SIMEX, and MLE estimators have the desirable property of being independent of mean prior achievement, making them more fair and appealing educator performance indicators.

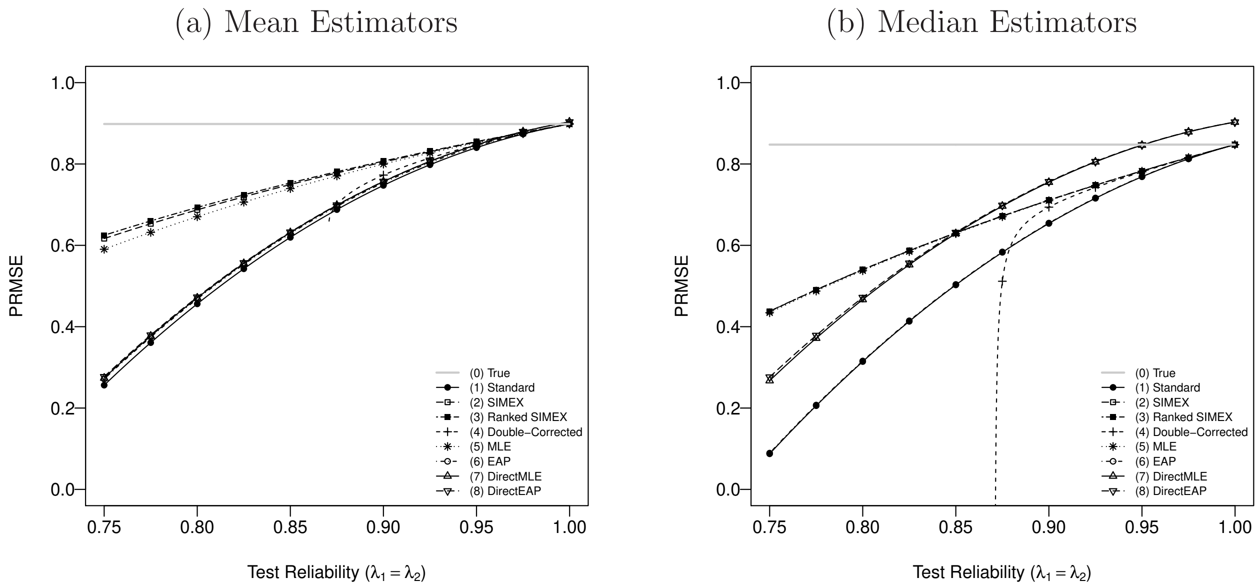

Figure B1: Comparing the PRMSE of MGP as a function of test reliability ($\lambda_1 = \lambda_2$) for (a) mean estimators and (b) median estimators.

The comparison also clearly demonstrates that meanGP have substantially higher marginal reliability than medGP. In terms of PRMSE, even the best performing medGP performs only slightly better than the worst performing meanGP. Among the estimators that aggregate individual student SGP, the means have PRMSEs of .14 to .20 higher than the medians at a test reliability of 0.9. These are substantial differences, that are likely to make the meanGP preferable in practice.

References

Carroll, R., Ruppert, D., Stefanski, L., & Crainiceanu, C. (2006). *Measurement error in nonlinear models: A modern perspective (2nd ed.)*. London: Chapman and Hall.

Lockwood, J. R., & Castellano, K. E. (2015). Alternative statistical frameworks for student growth percentile estimation. *Statistics and Public Policy*, *2*(1). doi:10.1080/2330443X.2014.962718

McCaffrey, D. F., Castellano, K. E., & Lockwood, J. R. (2015). The impact of measurement error on the accuracy of individual and aggregate SGP. *Educational Measurement: Issues and Practice*, *34*, 15–21. doi:10.1111/emip.12062

Monroe, S., & Cai, L. (2015). Examining the reliability of student growth percentiles using multidimensional IRT. *Educational Measurement: Issues and Practice*, *34*, 21–30. doi:10.1111/emip.12092

Shang, Y., Van Iwaarden, A., & Betebenner, D. (2015). Covariate measurement error correction for student growth percentiles using the SIMEX method. *Educational Measurement: Issues and Practice*, *34*, 4–14. doi:10.1111/emip.12058

Zacks, S. (1981). *Parametric statistical inference*. Oxford: Pergamon Press.