a not-for-profit organization

AUGUST 2026

The Mobility Penalty in School Value-added Accountability: Evidence from Ohio

Ian Kingsbury, Senior Fellow

Abstract

Value-added models (VAMs) are widely used in education accountability systems to estimate school and teacher effectiveness. However, myriad methodological concerns have been raised about the ability of these models to measure quality. Among those concerns is the risk that these systems may inadvertently penalize schools serving highly mobile student populations, as mobility is known to be detrimental to student learning.

This study examines school-level value-added data from Ohio to test whether higher student mobility rates are associated with lower VAM scores. Using publicly available datasets linking mobility metrics with value-added estimates, I find a robust negative relationship. That is, schools with elevated mobility exhibit systematically lower value-added measures, even after accounting for student demographics and baseline achievement.

Background

Value-added models (VAMs) have become a cornerstone of modern education accountability systems, promising to isolate as best as possible the contributions of schools and teachers to student learning growth by controlling for prior achievement and other factors. These models represent a shift from proficiency metrics in recognition of the fact that proficiency rates largely reflect student demographics and prior achievement rather than the quality of instruction provided by schools and teachers (Betebenner, 2012; Ho, 2008; Koedel et al., 2015; Polikoff, 2017; Polikoff et al., 2013). By focusing on growth, VAMs aim to provide a fairer and more accurate assessment of educational effectiveness.

The conceptual roots of value-added analysis in education can be traced to early production function studies in economics. Hanushek (1971) and Murnane (1975) explored teacher effects on student outcomes using regression-based approaches that accounted for prior performance. Broader interest surged in the 1990s with the work of William

Sanders and colleagues in Tennessee. Sanders and Rivers (1996) analyzed longitudinal student test score data and demonstrated that teacher effects are cumulative and persistent, with students assigned to ineffective teachers for multiple years showing lasting deficits even when later taught by effective instructors. This was followed by Wright, Horn, and Sanders (1997), who argued that teachers are the most significant source of variation in student achievement. These findings animated efforts to produce the Tennessee Value-Added Assessment System (TVAAS), which used multilevel modeling to estimate learning effects.

The No Child Left Behind Act (NCLB, 2001/2002) accelerated adoption by emphasizing test-based accountability, though it initially focused on proficiency measures. The subsequent Race to the Top initiative (2009) further incentivized states to incorporate growth or value-added metrics into teacher and school evaluations, often tying them to high-stakes decisions like tenure, compensation, and school ratings (Collins & Amrein-Beardsley, 2014; Shen et al, 2016; Özek & Xu, 2019; Thomas et al., 2019). By the 2010s, numerous states and districts had implemented VAMs (Pivovarova et al., 2016).

Ohio was an early adopter of value-added measures in a performance accountability framework. House Bill 3 (2003) required the state to incorporate a value-added progress dimension into district and school ratings to complement performance indices, with statewide report-card use following several years later (Lloyd, 2008). Ohio utilized the SAS Educational Value-Added Assessment System (EVAAS), rooted in Sanders' methodology (Lloyd, 2008). This allowed for estimates of student progress from one year to the next relative to expected growth based on prior achievement trajectories. Value-added became a key component of Ohio's school report cards and educator evaluations, providing estimates on how schools and teachers contribute to academic gains across subjects and grades. Critically, unlike Tennessee, Ohio also publishes school-level mobility data, making it an ideal case study to assess the relationship between mobility and value add measures.

Despite their widespread adoption and intuitive appeal, value-added models (VAMs) have faced substantial methodological scrutiny. These concerns are particularly acute when

VAMs are used for high-stakes accountability purposes, such as school ratings, teacher evaluations, or resource allocation (Koedel et al., 2015).

One core set of concerns revolves around bias in VAM estimates. Standard VAMs assume that, conditional on prior achievement and included covariates (e.g., student demographics), student assignment to schools or teachers is effectively random or that unobserved confounders are negligible. In reality, non-random sorting can introduce systematic bias. For instance, when it comes to teachers, Rothstein (2009, 2010) demonstrated that future teachers appear to "predict" prior student performance in ways that suggest sorting on unobservables, likely driven by parental preferences, administrative decisions, or unmeasured student characteristics. When it comes to schools, Paul and Greene (2022) analyzed student level data from virtual schools that provided reasons for enrollment. They observed that negative selection factors were predictive of lower ELA achievement even after controlling for prior achievement history and standard demographic data (e.g. race, gender, free or reduced lunch eligibility, special education status), an observation that casts doubt on the idea that the covariates in standard value add models are sufficient for producing unbiased estimates.

Even the strongest proponents of VAM would acknowledge that model specification matters greatly (and that it is not unreasonable to wonder or doubt whether the EVAAS system appropriately handles concerns around mobility). Different functional forms, choices of lagged scores (one versus multiple years), inclusion of student or peer covariates, and handling of missing data or attrition can produce meaningfully different rankings. The American Statistical Association (ASA, 2014) emphasized this sensitivity, noting that VAM scores and rankings can shift substantially under alternative specifications. Similarly, the inclusion (or exclusion) of sociodemographic adjustments has been debated on both technical and ideological grounds; omitting them may exacerbate bias in high-poverty settings, while including them raises questions about holding schools accountable only for factors within their control (Parsons et al., 2019).

Reliability and stability represent another major limitation. Teacher or school VAM estimates often exhibit low year-to-year correlations (frequently in the 0.2–0.5 range), raising doubts about their suitability for high-stakes decisions (Loeb & Candelaria, 2012; McCaffrey et al., 2009). This instability leads to high misclassification rates. Teachers or schools may be labeled "ineffective" one year and "effective" the next purely due to noise. Shrinkage estimators and multi-year averaging are common mitigations, but they do not fully resolve the issue, particularly for students with a history of low achievement (Herrmann et al., 2016).

Broader critiques, including those from the American Educational Research Association (AERA, 2015), stress that VAMs primarily capture correlations rather than causation and rely on standardized test scores that may not fully reflect teacher or school contributions to broader student outcomes (e.g., socioemotional development, long-term attainment, lifetime earnings).

Student mobility stands out as a particularly salient methodological concern in this literature, especially relevant to the current study. Highly mobile students disrupt the assumptions underlying standard VAMs (Cassiday, 2023). Mobility interrupts learning (Mehana & Reynolds, 2004), introduces gaps in student data histories (SAS Institute, 2013), complicates attribution of growth to a specific school (Cassiday, 2023), and correlates temporally with other potentially harmful disruptions (Pribesh & Downey, 1999). VAMs can bias estimates downward for schools serving transient populations, as these schools may inherit students with disrupted learning trajectories while receiving limited credit for prior contributions from "sending" schools (Goldstein et al., 2007).

Research consistently documents negative associations between mobility and student achievement (Mehana & Reynolds, 2004; Hanushek et al., 2004; Grigg, 2012; Pribesh & Downey, 1999). Mobile students often experience learning losses equivalent to several months of instruction, with effects compounding for frequent moves (Goldhaber et al., 2022). Within-year mobility appears especially disruptive (Min, 2022; Prior & Leckie, 2022; Guarino et al., 2025). These effects are not purely compositional; even after controlling for student characteristics, mobility predicts lower test scores and growth (Grigg, 2012; Paul & Greene, 2022). In VAM contexts, this hypothetically creates a structural disadvantage for high-mobility schools since they serve more students with unobserved shocks (e.g., family stress, adjustment periods), leading to systematically lower value-added scores.

Empirical work reinforces these concerns. Student moves reduce achievement growth, and high school-level turnover harms movers and non-movers alike, with larger costs in schools serving lower-income and minority students (Hanushek, Kain, & Rivkin, 2004). This pattern risks distorting accountability by penalizing schools for factors at least partially outside their control, such as student churn driven by external socioeconomic forces. Plausible outcomes that should concern policymakers are that these empirical realities could discourage schools from enrolling mobile students (Weckstein, 2003; Rhodes, 2005; Özek, 2012) or penalize schools that are serving students well but register low VAM scores.

In summary, while VAMs advanced accountability beyond static proficiency rates by emphasizing growth, they remain a technically and philosophically imperfect approach for measuring school quality. Schools in high-mobility environments may appear less effective not due to inferior instruction but because models imperfectly isolate school contributions. These issues motivate the present study's empirical examination of mobility-VAM relationships in Ohio data.

Data and Methods

This study uses publicly available school-level data from the Ohio Department of Education and Workforce's School Report Cards. The primary outcome is the Overall Effect Size from the 2024–25 Progress component. Ohio produces its value-added measures through the SAS Education Value-Added Assessment System (EVAAS). EVAAS estimates each student's expected academic growth based on their prior test-score trajectory and then compares actual performance to that expectation. The school-level Effect Size aggregates these student-level growth residuals into a single standardized metric of the school's contribution to learning. Positive values indicate that students, on average, grew more than expected given their prior achievement; negative values indicate below-expected growth; and a value of zero indicates growth exactly in line with expectations. For illustration, an effect size of +0.10 means students grew roughly one-tenth of a standard deviation more than expected, while an effect size of –0.20 indicates growth about one-fifth of a standard deviation below expectations.

Student mobility is measured as the percentage of students who entered or left the school during the 2024–25 school year. The analysis examines both the continuous mobility rate and a binary indicator for high-mobility schools, defined as a mobility rate of 25 percent or higher—the threshold used by the state of Ohio. Additional school-level covariates include contemporary (2024–25) measures of the natural log of enrollment, attendance rate, chronic absenteeism rate, and the percentage of students identified as economically disadvantaged, an indicator based primarily on free- or reduced-price lunch eligibility and related income criteria. Subsequent specifications add school racial composition (percent Black, Hispanic, and White) and lagged school-level proficiency (the share of students scoring at or above Proficient on 2023–24 assessments). Inclusion of race/ethnicity as covariates is limited to those three since data is suppressed at n<10. Including those three variables already results in losing nearly half the sample, while additional controls (e.g. Asian, Native American, multiracial) would decrease the number substantially further still.

The analytic sample is restricted to schools with non-missing values on the effect size and mobility rate, yielding 3,005 observations in the most inclusive models. Sample sizes decline when race and prior-proficiency variables are added because of data suppression rules. Models are estimated using ordinary least squares with robust standard errors. Specifications are built sequentially, beginning with a bivariate association and then adding covariates in stages.

This sequencing is deliberate. Enrollment is the most straightforward control and poses little risk of over-control. The remaining covariates are more ambiguous. Attendance and chronic absenteeism may be both causes and consequences of mobility and school climate. Economic disadvantage and racial composition are correlated with mobility and with the unobserved shocks that high-mobility schools absorb. Lagged school-level proficiency is especially fraught. Because the outcome is already a student-level growth residual, further controlling for aggregate prior achievement can partial out part of the very disadvantage the analysis is designed to detect. These variables are therefore presented as robustness checks rather than as a single preferred specification. The central empirical question is whether the mobility association survives this sequence of increasingly conservative models.

Results and Discussion

The analysis estimates the association between student mobility and school value-added using ordinary least squares. The continuous-mobility specification takes the form

$$Y_s = \beta_0 + \beta_1 \text{Mobility}_s + X_s \gamma + \varepsilon_s$$

where Y_s is the Overall Effect Size for school s, Mobility_s is the continuous mobility rate, and X_s is a vector of school-level covariates that expands across columns. The binary high-mobility models replace the continuous rate with an indicator equal to one if the mobility rate is 25 percent or higher:

$$Y_s = \beta_0 + \beta_1 \text{HighMobility}_s + X_s \gamma + \varepsilon_s.$$

Figure One highlights a clear, negative relationship between student mobility and value-added effect sizes.

Figure One: Relationship between mobility and value-added scores

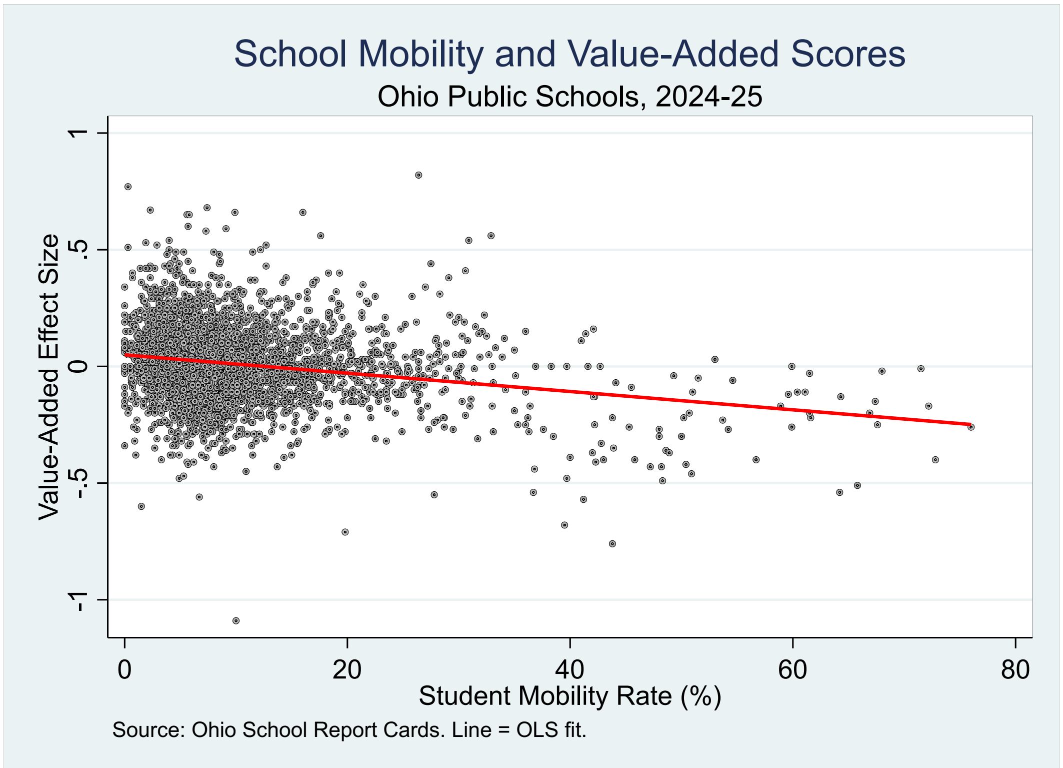

Table 1 reports results from the continuous specification and clarifies that within the bivariate model (Column I), a one-percentage-point increase in the mobility rate is associated with a 0.0039-point decline in effect size ($p < .01$). Adding enrollment (Column II) and then attendance, chronic absenteeism, and economic disadvantage (Column III) reduces the magnitude but leaves the coefficient negative and statistically significant ($p < .01$). Adding racial composition (Column IV) cuts the sample size nearly in half but the magnitude increases to -0.0052 and remains statistically significant ($p < .01$). Including lagged school-level proficiency (Column V) attenuates the estimate to –0.0023, but the coefficient nonetheless remains negative and statistically significant ($p < .01$).

Table One: Predictors of school value-added effect size (linear mobility measure)

	I	II	III	IV	V
Mobility	-0.0039*** (0.0004)	-0.0035*** (0.0004)	-0.0031*** (0.0007)	-0.0052*** (0.0007)	- 0.0023*** (0.0006)
Ln(Enrollment)	-	0.0275*** (0.0050)	0.0259 (0.0049)***	0.0037 (0.0077)	0.0019 (0.0071)
Attendance	-	-	0.002* (0.0010)	0.0001 (.0013)	0.0039*** (0.0012)
Chronic Absenteeism	-	-	0.0005 (0.0004)	-0.0007 (0.0006)	0.0024*** (0.0006)
Econ Disadvantage	-	-	-0.0001 (0.0001)	-0.0006*** (0.0002)	0.0001 (0.0002)
Black	-	-	-	0.0003 (0.0006)	0.0010* (0.0006)
Hispanic	-	-	-	0.0009 (0.0007)	0.0019*** (0.0006)
White	-	-	-	-0.0009 (0.0006)	-0.0009 (0.0006)
Prior proficiency	-	-	-	-	0.0045*** (0.0004)
n	3,005	3,004	2,999	1,533	1,529

Ohio defines a school as “high mobility” if at least 25 percent of its enrollment consists of students who have attended the school for less than one year (Ohio Rev. Code § 3302.03). Following that 25 percent threshold, additional models treat mobility as a binary high-mobility indicator rather than as a continuous rate. This specification is useful if the relationship is not strictly linear and perhaps does not manifest below certain mobility thresholds.

Highly mobile schools have effect sizes 0.1078 points lower than other schools in the bivariate model ($p < .01$). The gap narrows with additional controls but remains statistically significant in every specification, including the fullest model that includes race and prior proficiency (–0.047, $p < .01$).

Table Two: Predictors of school value-added effect size (high mobility schools)

	I	II	III	IV	V
Highly mobile	-0.1078*** (.0173)	-0.0917*** (0.0180)	-0.052** (0.0228)	-0.0958*** (0.0251)	- 0.0467*** (0.0161)
Ln(Enrollment)	-	0.032 (0.0051)	0.0273*** (0.0050)	0.0067 (0.0081)	0.0026 (0.0071)
Attendance	-	-	-0.0003** (0.0001)	0.0022* (0.0012)	0.0048*** (0.0011)
Chronic Absenteeism	-	-	0.0031*** (0.0011)	-0.0008 (0.0006)	0.0026*** (0.0006)
Econ Disadvantage	-	-	0.0003 (0.0004)	-0.0008*** (0.0002)	0.0001 (0.0002)
Black	-	-	-	0.0004 (0.0007)	0.0011* (0.0006)
Hispanic	-	-	-	0.0010 (0.0007)	0.0020*** (0.0007)
White	-	-	-	-0.0006 (0.0006)	-0.0007 (0.0006)
Prior Proficiency	-	-	-	-	0.0049*** (0.0004)
n	3,005	3,004	2,999	1,533	1,529

Descriptive patterns in Table 3 illustrate the practical magnitude of these associations. Across the full sample of 3,005 schools, the mean effect size is 0.010 and 54.3 percent of schools post non-negative scores. The disadvantage widens at higher mobility thresholds: schools above 30 percent mobility average –0.147 (28.6 percent non-negative), and those above 40 percent average –0.226 (only 11.7 percent non-negative). High-mobility schools are concentrated in the lower tail of the value-added distribution.

Table Three: Effect size outcomes by mobility bucket

Mobility Rate	*N Schools*	*Mean Effect Size*	*Median Effect Size*	*% with Effect Size ≥ 0*
All schools	3,005	0.010	0.00	54.3%
> 10%	1,021	–0.017	–0.02	46.1%
> 20%	299	–0.059	–0.04	38.8%
> 30%	112	–0.147	–0.15	28.6%
> 40%	60	–0.226	–0.22	11.7%

It bears repeating that chronic absenteeism, economic disadvantage, and racial composition are entangled with the conditions that produce mobility and with the unobserved shocks that mobile students bring into the growth calculation. Prior proficiency is an especially aggressive control because the dependent variable is already residualized on student prior achievement, so adding school-level lagged proficiency can absorb part of the mobility penalty rather than isolate confounding. The attenuation observed when controlling for prior achievement is therefore expected and should not be read as evidence that the association between mobility and VAM is spurious with sufficient controls in place. Indeed, it is notable that the negative mobility coefficient survives every specification. That persistence, including in the most conservative models, is the result that deserves attention.

While reverse causality (i.e., low value-added scores prompting parental exit) cannot be ruled out entirely, extant literature provides some reassurance. Research on school choice behavior, including analyses of New York City high school applications (Abdulkadiroğlu et al., 2020), finds that parents respond primarily to absolute achievement levels and peer composition rather than to growth or value-added measures. Because the models already control for economic disadvantage, attendance, racial composition, and (in the fullest specifications) prior proficiency, the prospect of reverse causality operating through VAM-driven exit is reduced.

The robustness and magnitude of the mobility–value-added relationship have direct implications for accountability. Ohio uses value-added results to construct the Progress Component of school report cards. Progress accounts for 25 percent of a school's overall report-card rating or a slightly larger share when some components are missing. Progress scores are a function of both the growth index (statistical certainty) and the effect size (magnitude of growth). For schools, a 5-star Progress rating requires an effect size of at least +0.20 while a 1-star rating requires a score of -0.20 or less. As a result, 32.2% of highly mobile schools in the sample have effect sizes that qualify them for a 1-star rating compared to 7.3% of non-highly mobile schools.

In sum, higher student mobility is robustly associated with lower value-added scores in Ohio. The relationship holds after accounting for enrollment, attendance, chronic absenteeism, economic disadvantage, racial composition, and lagged proficiency, and it is most pronounced among schools meeting or exceeding the state's own 25 percent high-mobility threshold. While the data cannot isolate a purely causal effect, the fact that the negative coefficient is maintained in all model specifications is a signal that current growth measures imperfectly isolate school contributions in high-mobility settings.

Notably, Tennessee, North Carolina, and Pennsylvania also use EVAAS value-added measures in statewide accountability. Those states do not publish statewide school-level mobility rates comparable to Ohio's, but there is little reason to think the patterns documented here are unique to Ohio. In those states as in Ohio, policymakers seeking fairer accountability may wish to consider solutions including but not limited to mobility-adjusted growth metrics, complementary indicators of student stability, or reduced weight on value-added results for schools that serve highly mobile populations.

References

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., & Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review, 110*(5), 1502–1539. https://doi.org/10.1257/aer.20172040

American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44*(8), 448–452.

American Statistical Association. (2014). ASA statement on using value-added models for educational assessment. https://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf

Betebenner, D. W. (2012). Growth, standards, and accountability. In G. J. Cizek (Ed.), Setting performance standards: Foundations, methods, and innovations (2nd ed., pp. 439–450). Routledge.

Cassiday, K. R. (2023). Accounting for student mobility in school rankings: A comparison of estimates from value-added and multiple membership models (Doctoral dissertation, University of Maryland).

Collins, C., & Amrein-Beardsley, A. (2014). Putting Growth and Value-Added Models on the Map: A National Overview. *Teachers College Record, 116*(1).

Goldhaber, D., Koedel, C., Özek, U., & Parsons, E. (2022). Using longitudinal student mobility to identify at-risk students. *AERA Open, 8*. https://doi.org/10.1177/23328584211071090

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(4), 941–954.

Grigg, J. (2012). School enrollment changes and student achievement growth: A case study in an urban school district. *Sociology of Education, 85*(4), 388–404

Guarino, C., Li, Y., Santibañez, L., Kang, H., & Ream, R. K. (2025). What is the impact of changing schools on the academic outcomes of elementary and middle school students? (EdWorkingPaper No. 25-1318). Annenberg Institute at Brown University. https://doi.org/10.26300/w73j-x228

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review, 61*(2), 280–288.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Disruption versus Tiebout improvement: The costs and benefits of switching schools. *Journal of Public Economics, 88*(9–10), 1721–1746.

Herrmann, M., Walsh, E., & Isenberg, E. (2016). Shrinkage of value-added estimates and characteristics of students with hard-to-predict achievement levels. Statistics and Public Policy, 3(1), 1–10

Ho, A. D. (2008). The problem with “proficiency”: Limitations of statistics and policy under No Child Left Behind. *Educational Researcher, 37*(6), 351–360.

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

Lloyd, J. L. (2008). Instituting a value added assessment system in Ohio: An investigation of the factors that influenced the decision [Doctoral dissertation, Ashland University]. OhioLINK Electronic Theses and Dissertations Center. https://etd.ohiolink.edu/acprod/odb_etd/ws/send_file/send?accession=ashland1226675998&disposition=inline

Loeb, S., & Candelaria, C. A. (2012). How stable are value-added estimates across years, subjects, and student groups? Carnegie Knowledge Network. http://www.carnegieknowledgenetwork.org/briefs/value-added/value-added-stability/

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal stability of teacher effect estimates. *Education Finance and Policy, 4*(2), 164–187.

Mehana, M., & Reynolds, A. J. (2004). School mobility and achievement: A meta-analysis. *Children and Youth Services Review, 26*(1), 93–119.

Min, J. (2022). Between-year and within-year school mobility: Different effects by race/ethnicity. *Education and Urban Society, 54*(8), 939–961. https://doi.org/10.1177/00131245211004563

Murnane, R. J. (1975). The impact of school resources on the learning of inner city children. Ballinger.

Ohio General Assembly. (2003). *Am. Sub. H.B. 3, 125th General Assembly*. https://search-prod.lis.state.oh.us/api/v2/general_assembly_125/legislation/hb3/04_PS/pdf/

Ohio Rev. Code § 3302.03 (2025). https://codes.ohio.gov/ohio-revised-code/section-3302.03

Özek, U. (2012). *One day too late? Mobile students in an era of accountability* (CALDER Working Paper No. 82). National Center for Analysis of Longitudinal Data in Education Research. https://caldercenter.org/publications/one-day-too-late-mobile-students-era-accountability

Özek, U., & Xu, Z. (2019). Misattribution of teacher value added. *Education Finance and Policy, 14*(1), 115–148.

Parsons, E., Koedel, C., & Tan, L. (2019). Accounting for student disadvantage in value-added models. *Journal of Educational and Behavioral Statistics, 44*(2), 144–179

Paul, J. D., & Greene, J. P. (2022). *Investigating the relationship between negative selection into online schooling and achievement growth* (EDRE Working Paper No. 2022-02). University of Arkansas. https://scholarworks.uark.edu/edrepub/133/

Pivovarova, M., Amrein-Beardsley, A., & Broatch, J. (2016). Value-added models (VAMs): Caveat emptor. *Statistics and Public Policy, 3*(1), 1–9

Polikoff, M. (2017, March 20). Proficiency vs. growth: Toward a better measure. *FutureEd*. https://www.future-ed.org/proficiency-vs-growth-toward-a-better-measure/

Polikoff, M. S., & Wrabel, S. L. (2013). When is growth not growth? The use of growth percentiles in school accountability systems. *Educational Researcher, 42*(5), 259–269.

Pribesh, S., & Downey, D. B. (1999). Why are residential and school moves associated with poor school performance? *Demography, 36*(4), 521–534.

Prior, L., & Leckie, G. (2022). Student mobility: Extent, impacts and predictors of a range of movement types for secondary school students in England. *British Educational Research Journal, 48*(5), 1027–1055. https://doi.org/10.1002/berj.3807

Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy, 4*(4), 492–519.

Rhodes, V. L. (2005). Kids on the move: The effects of student mobility on NCLB school accountability ratings. *Perspectives on Urban Education, 3*(3).

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

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. The Quarterly Journal of Economics, 125(1), 175–214.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. University of Tennessee Value-Added Research and Assessment Center.

SAS Institute. (2013, May 17). Value-added myth busting, Part 4: Value-added models cannot measure growth of students who have missing data or are highly mobile. SAS Blogs.

Shen, Z., Simon, C. E., & Kelcey, B. (2016). The potential consequence of using value-added models to evaluate teachers. *eJEP: eJournal of Education Policy*, Fall 2016.

Thomas, D., & Wieczorek, D. (2019). What did we learn from Race to the Top teacher evaluation systems? *AASA Journal of Scholarship & Practice, 16*(2), 18–34.

Weckstein, P. (2003). Accountability and student mobility under Title I of the No Child Left Behind Act. *Journal of Negro Education, 72*(1), 117–125.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57–67.