Selected Statistical Models for Incorporating Student Academic Growth in Estimating Effect of Teachers in Tested Grades and Subjects: Technical Comparison Matrix

The expert panel appointed by the North Carolina Department of Public Instruction (DPI) has rated the following three value-added models (VAMs) as having the technical adequacy needed for the high-stakes purpose of estimating teacher effectiveness: a three-level hierarchical linear model (HLM III); a univariate Education Value-Added Assessment system (EVAAS) model developed by William Sanders (EVAAS-URM); and a student fixed effects model (SFE). WestEd was asked to assist the DPI by reviewing the findings from the expert panel and independently validating the appropriateness of each model for the intended purpose.

As a first step, WestEd reviewed existing literature on various statistical models that are commonly used to estimate the effect of teachers on students' academic growth. This step was intended to confirm the expert panel's decision to look only at VAMs in its analyses. Then, following the release of the draft technical report from the expert panel, we focused our attention specifically on the three panel-identified models. Our findings are synthesized in the following matrix.

Note: We also included a multivariate EVAAS (EVAAS-MRM) in our review. Though this model was not one of the three recommended by the panel, our expert review of the literature suggested that this model has unique features that the DPI may want to consider during decision-making.

Model	Key Features	Strengths	Limitations	Considerations	References
Three-Level Hierarchical Linear Model (HLM III)	Multilevel analyses take nested nature of data (i.e., students nested within teachers and teachers within schools) into account Random effects model: teacher effect is residual after student characteristics (previous year's score + measured background characteristics), measured classroom characteristics, and measured school characteristics are accounted for Describes relative contributions of school, teacher, and student to current student performance Growth curves for each student are constructed to show the shaping	Has capacity to include/account for multiple sources of variation, nesting students within classrooms (teachers) within schools, ideally allowing for more precise estimates Student-, teacher, and school-level covariates (e.g., SES) can be added Minimizes measurement error by simultaneously modeling within- and between-school variance Takes advantage of all available information but can conduct with only two data points (test scores) for each student Due to the particular parameters for	Requires moderately high computing resources relative to other models considered Does not account for teachers' effect on student progress over time (e.g., extending into years beyond the current year taught) Cannot account for or correlate data for teachers who teach across grades and subjects Expert panel found it least consistent in identifying top/ bottom 5% using actual data	Schochet and Chiang (2010) found that in order to ensure accuracy in estimating teacher effects, test scores must be available for all students, assessments must be vertically scaled across grades, and all teachers must be stationary (school and grade) for all years in which data are collected Unbiased estimates are more realistic if one is making within-school comparisons rather than between-school comparisons Because there are three	Schochet & Chiang (2010) Raudenbush & Bryk (2002) Webster (2005)

Model	Key Features	Strengths	Limitations	Considerations	References
	effects of individual and contextual factors on academic progress over time Tracks a teacher's annual effect, not cumulative effects over time Does not require that tests are vertically scaled across grades	specification, model does not lose precision in identifying effective teachers for smaller class sizes Statistical adjustments possible to allow for missing data		levels, requiring three separate specifications of parameters, the potential for misspecification (and therefore error) is greater than for two-level models Dallas ISD prefers two-level HLM over three-level (Webster, 2005)	
Univariate Response Model (EVAAS-URM)	Random effects model: teacher effect is residual after student's previous test score is accounted for; no other student-, teacher-, or school-level characteristics are included Model is similar to ANCOVA in that students' previous scores are believed to be associated with their scores in subsequent years, and therefore, the previous scores are included in the model as covariates Includes any student who has at least three data points (test scores) Can be aggregated to report school, district, or statewide effect Tracks a teacher's annual effect, not cumulative effects over time Does not require that assessments yield scores that are on the same scale (i.e., vertically linked)	Uses all data for each student if at least three prior test scores are available Statistical adjustments possible to allow for missing data Attempts to mitigate the effects of nonrandom assignment of teachers to schools For classroom-level analysis, accommodates team teaching, departmentalized instruction, and self-contained classrooms Can be used to project future student achievement outcomes Yields conservative estimates of teacher effectiveness, intended to reduce risk of misidentification	Requires moderately high computing resources relative to other models considered Does not explicitly account for past teachers' effect on student growth This EVAAS model is less widely used than EVAAS-MRM, so less research-based support was found	The URM is the best model when data structures do not meet the requirements for an MRM analysis (Sanders, 2010 Students act as their own controls; therefore, SES and demographic factors do not need to be factored in Student-level covariates are not added, as these characteristics are assumed to be fixed; does not make adjustments for students' demographic characteristics	Sanders (2006) Sanders (2010) Sanders & Wright (2009) Sanders, Wright, Rivers, & Leandro (2009)

Model	Key Features	Strengths	Limitations	Considerations	References
Student Fixed Effects Model (SFE)	Teacher effect is mean of residuals for all students (aggregate) after accounting for students' prior performances; student effect (including ability) is "fixed"—i.e., it does not vary across teachers (hence, between-student variance is controlled, as each student serves as his or her own control) Tracks a teacher's annual effect, not cumulative effects over time Conditioned only on student's prior year's performance Does not require that assessments yield scores that are on the same scale (i.e., vertically linked)	Statistically parsimonious and can be conducted on commonly available software Found to be as accurate as the EVAAS-MRM in estimating growth, and is far simpler and less expensive to implement (Tekwe et al., 2004) Found to provide consistent estimates (Harris & Sass, 2006) Used in recent study that detected a persistent teacher effect (Chetty, Friedman, & Rockoff, 2011)	Does not explicitly include student-, teacher-, or school-level demographic factors Does not account for multiple schools attended (if such was the case) or the cumulative effect of past or future gains of teaching Does not produce shrunken estimates, which are considered to be a theoretically superior method of estimating value-added effects Did not perform as strongly as other VAMs (Kane & Staiger, 2008)	Can be viewed as a feature of a model rather than as a unique model Consistency of teacher estimates depends on teacher stability in grade taught and home school Introducing fixed effects to VAMs in an effort to compensate for nonrandom assignment of teachers to classrooms is unwarranted (Rothstein, 2007) Student-level covariates are not added, as these characteristics are assumed to be fixed	Chetty, Friedman, & Rockoff (2011) Harris & Sass (2006) Clotfelter, Ladd, & Vigdor (2007) Rothstein (2007) Kane & Staiger (2008) Tekwe, Carter, Ma, Algina, & Lucas (2004)
Multivariate Response Model (EVAAS-MRM)	A multivariate, longitudinal, linear mixed model (multivariate repeated-measures ANOVA model) Accounts for students clustering with teachers and for students and peers clustering with different teachers in different years and with multiple teachers in a given year in the same subject The entire set of observed test scores belonging to each student is fitted simultaneously Teacher effects are treated as random—obtained through	Uses all past, present, and future data available for each student Statistical adjustments possible to allow for missing data Can account for the possibility that a teacher is effective for one subject and ineffective for another Allows teacher effects to be accumulated over time Accommodates team teaching; allows for fractional assignment of effect (Sanders, Wright, Rivers, & Leandro, 2009)	Most statistically complex, relative to other models considered; uses proprietary software Because effectiveness is reported in terms of gains, assessments must be vertically scaled Works well with end-of-grade tests given in elementary and middle school, in which each annual prior-grade assessment serves as a data point; more challenging with high school end-of-course tests that do not have a built-in pre-test Does not include student- or	Best used when tests are vertically scaled, to allow comparable expectations of progress—evaluated over many schools and/or districts—regardless of entering levels of groups of students Students act as their own controls; therefore, socioeconomic and demographic factors do not need to be factored in	Sanders (2006) Sanders (2010) Sanders, Wright, Rivers, & Leandro (2009) Schochet & Chiang (2010)

Model	Key Features	Strengths	Limitations	Considerations	References
	shrinkage estimation; this means that a teacher is considered to be "average" (with a teacher effect of zero) until there is sufficient student data to indicate otherwise Does not require that assessments yield scores that are on the same scale (i.e., vertically linked)	Provides conservative estimates of teacher effects so that effective teachers are unlikely to be misclassified as ineffective Guards against random influences on student achievement that are not attributable to educational intent Attempts to mitigate the effects of nonrandom assignment of teachers to schools This original EVAAS model has been widely researched; strengths and limitations are well documented	classroom-level covariates, so teachers are held equally accountable for progress of all students Assumes that teacher effects persist undiminished over time and are additive		

References

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects.* Durham, NC: Sanford Institute of Public Policy, Duke University.

Harris, D., & Sass, T. R. (2006). *Value-added models and the measurement of teacher quality*. Washington, DC: Institution for Education Sciences.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). London: Sage.

Rothstein, J. (2007). *Do value-added models add value? Tracking, fixed effects, and causal inference*. Cambridge, MA: National Bureau for Economic Research.

Sanders, W. L. (2006, October). *Comparisons among various educational assessment value added models*. Paper presented at The Power of Two—National Value-Added Conference.

Sanders, W. L. (2010). *EVAAS statistical models*. SAS White Paper. Cary, NC: SAS.

Sanders, W. L., & Wright, S. P. (2009). *A response to Amrein-Beardsley (2008) "Methodological concerns about the education value-added assessment system*." SAS White Paper. Cary, NC: SAS.

Sanders, W. L., Wright, S. P., Rivers, J. C., & Leandro, J. G. (2009). *A response to criticisms of SAS EVAAS*. SAS White Paper. Cary, NC: SAS.

Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains.* Washington , DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Tekwe, C. D., Carter, R. L., Ma, C. X., Algina, J., & Lucas, M. E. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29(1), 11–36.

Webster, W. J. (2005). The Dallas school-level accountability model: The marriage of status and value-added approaches. In R. W. Lissitz (Ed.), *Value added models in education: Theory and applications* (pp. 233–271). Maple Grove, MN: JAM Press.