Precision of Measures of Central Tendency: Computing an Effectiveness Index for Teachers

Dash Weerasinghe
Mark Anderson
Karen Bembry
Dallas Public Schools

Abstract

In an age of student accountability, public school systems must find procedures for identifying teachers who help students continue to grow academically. It is not sufficient to identify effective campuses; instead accountability must be taken down to the classroom level. The final component in a district accountability system should be procedures for identifying effective classrooms. However, the effectiveness measure must be consistent and stable. This paper describes a study investigating measures of central tendency and adjustments to those measures to determine which combination is the most precise. It also establishes the minimum number of subjects needed to calculate a reliable measure.

The development of a value-added student achievement model for individual classroom teachers evolved from Dallas Public School’s efforts to equitably measure school and classroom effects in order to hold schools, principals, and teachers accountable for growth in student achievement. Beginning in 1995, Classroom Effectiveness Indices (CEIs) have been calculated for each teacher in the district who administered an achievement test. Because the CEIs were for planning purposes only— they were to serve as an indicator for a teacher and an administrator of that teacher’s classwide improvement—CEIs were generated for all classrooms regardless of the number of students computed into the overall CEI. Because CEIs are used primarily for planning purposes, with no consequences attached, the stability of the CEIs or the

minimum number of student residuals needed for a reliable measure have not been investigated.

**Purpose of this study:**

The purpose of this study is to investigate measures of central tendency and adjustments to the measures to determine which combination generates the most precise CEI. The minimum number of student residuals needed to calculate a reliable measure is also investigated.

**Data Source:**

Data for eighth grade teachers (n = 418) and data for fourth grade teachers (n = 592) comprised the Classroom Effectiveness Indices investigated. Stanford 9 test scores in reading and math were the outcome variables. Only continuously enrolled students were included. Continuously enrolled is defined as any student enrolled on a school campus by the first day of the second six-week grading period through the spring test date. Students included in this study had achievement scores from the previous school year. Students with excessive absences, defined as 20 or more per year, were removed prior to the final stage.

**Classroom Effectiveness Indices:**

Classroom Effectiveness Indices are measures of student achievement using two years of standardized test scores and other student-level and school-level covariates. For the purpose of this study, student scores on the Spring 2000 Stanford 9 in reading and mathematics subtests will be the outcome variables of achievement. Previous achievement used as the predictor variables will be the ITBS mathematics and reading scores from the previous school year. The standardized test scores are computed into
individual residualized gain scores using a linear regression and multi-level model, where student and school characteristics are regressed out from student gain scores. These residualized gains are grouped and assigned to teachers using student and teacher identification numbers, compiling a database of student gain scores by teacher.

Classroom Effectiveness Indices are computed in two stages using multiple regression-based procedures. In the first stage, outcome and predictor variables are regressed against covariates called fairness variables using multiple regression. Student test scores are regressed against nine student level characteristics or covariates.

Covariates include ethnicity, limited English proficiency status, gender, and variables indicating socio-economic status (SES).

$Y_{ij} = \text{Outcome variable of interest for each student } i \text{ in school } j.$

$X_{1j} = \text{African American English proficient student (1 if yes, 0 if all others)}$

$X_{2j} = \text{Hispanic English proficient (1 if yes, 0 if all others)}$

$X_{3j} = \text{Limited English Proficient (1 if yes, 0 if all others)}$

$X_{4j} = \text{Gender (1 if male, 0 if female)}$

$X_{5j} = \text{Free/reduced lunch (1 if yes, 0 if not on free/reduced lunch)}$

$X_{6j} = \text{Block-level average family income}$

$X_{7j} = \text{Block level average family education}$

$X_{8j} = \text{Block level average family poverty index}$

$X_{9j} = \text{variable } k \text{ for the } k^{th} \text{ student in school } j$
Stage 1 Regression Equation:

\[ Y_{ij} = \beta_0 + \beta_1X_{ij1} + \beta_2X_{ij2} + \beta_3X_{ij3} + \beta_4X_{ij4} + \beta_5X_{ij5} + \beta_6X_{ij6} + \beta_7X_{ij7} + \beta_8X_{ij8} + \beta_9(X_{ij9}X_{ij4}) + \beta_{10}(X_{ij9}X_{ij6}) + \beta_{11}(X_{ij9}X_{ij7}) + \beta_{12}(X_{ij9}X_{ij8}) + \beta_{13}(X_{ij9}X_{ij3}) + \beta_{14}(X_{ij9}X_{ij5}) + \beta_{15}(X_{ij9}X_{ij10}) + \beta_{16}(X_{ij9}X_{ij12}) + \beta_{17}(X_{ij9}X_{ij13}) + e_{ij} \]

Where \( e_{ij} \sim N(0, \sigma^2) \)

At the second level, a hierarchical model adjusts for school level variables as regressed against the residualized student level gain scores. The school level covariates are as follows:

\( W_{ij} = \) school mobility
\( W_{2i} = \) school overcrowdedness
\( W_{3i} = \) school average family education
\( W_{4i} = \) school average family income
\( W_{5i} = \) school average family poverty index
\( W_{6i} = \) school percent on free/reduced lunch
\( W_{7i} = \) school percent minority
\( W_{8i} = \) school percent African American
\( W_{9i} = \) school percent Hispanic
\( W_{10i} = \) school percent limited English proficient

The second stage equations:

\[ \beta_{kj} = \gamma_{k0} + \gamma_{k1}W_{1j} + \gamma_{k2}W_{2j} + \gamma_{k3}W_{3j} + \gamma_{k4}W_{4j} + \gamma_{k5}W_{5j} + \gamma_{k6}W_{6j} + \gamma_{k7}W_{7j} + \gamma_{k8}W_{8j} + \gamma_{k9}W_{9j} + \gamma_{k10}W_{10j} + u_{kj} \]

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \gamma_{03}W_{3j} + \gamma_{04}W_{4j} + \gamma_{05}W_{5j} + \gamma_{06}W_{6j} + \gamma_{07}W_{7j} + \gamma_{08}W_{8j} + \gamma_{09}W_{9j} + \gamma_{010}W_{10j} + u_{0j} \]
To calculate the CEI for classroom \( t \) in school \( j \) with \( K_t \) students:

\[
CEI_j = \frac{1}{K_t} \sum_{k=1}^{K_t} y_{kj}
\]

The CEI is calculated with respect to the district regression line. The CEI for teacher \( t \) in school \( j \) is then adjusted as follows:

\[
\text{Shrinkage Adjustment} = \frac{1}{1 + (\text{VAR} / \text{N})}
\]

Each student has a residualized gain that can be aggregated to a teacher identification number and a course number. All remaining residuals are standardized to a mean of 50 and a standard deviation of 10. Students with excessive absences are removed, and an overall Classroom Effectiveness Index is computed.

**Design**

For this analysis, all fourth grade and eighth grade teachers were included who had residualized gain scores for either Stanford 9 mathematics or reading. For each of the teachers, four measures of central tendency will be computed: the mean, a second mean with the highest 5% and the lowest 5% of the residualized scores removed (called a 90% trimmed mean), a third mean with the highest and lowest 10% of the residualized gain scores removed (called a 80% trimmed mean), and the median. The standard error for each corresponding measure will also be calculated. The standard error of the median will be estimated using bootstrapping procedures (\( n = 200 \)), where repeated sampling with replacement will be carried out to determine the variation. The teachers were grouped by the number of students in the index, and for each measure of that group the minimum and maximum standard errors were calculated.
Results:

The standard errors of the mean varied from 0.4342 - 0.0779 for 4th grade math and from 0.5936 - 0.0976 for 4th grade reading. The standard errors of the mean varied from 0.3111 - 0.0837 for 8th grade math and from 0.3939 - 0.0817 for 8th grade reading. There were several spikes in the standard errors as indicated on the accompanying charts. For example, the standard error for sections of 41 students is 0.1276, the standard error for sections of 42 students is 0.1947, and the standard error for sections of 43 students is 0.1319.

The 90% mean and the 80% mean have obviously lower standard error variations than the mean for most values of N. An example of the difference is 4th grade math for sections of 42 students, the SE of the mean is 0.1947, the SE of the 90% trimmed mean is 0.1716, and the SE of the 80% trimmed mean is 0.1559.

The median as a measure was the most consistent for most ranges of N, and the difference between the minimum and maximum standard error was less for the median than for other measures. The standard errors of the median ranged from 0.0778 - 0.0386 in 4th grade math and from 0.0418 - 0.0634 in 4th grade reading. The standard errors of the median on the 8th grade tests ranged from 0.0799 - 0.0379 in math and from 0.0271 - 0.0065 in reading. However, the median does not necessarily reflect outliers in low N that perhaps should be reflected in a Classroom Effectiveness Index.

The ideal solution to the problem is to use different measures of central tendency depending on the number of student residuals available for each teacher. The truncated means obviously had advantages over the mean overall in controlling for the influence of outliers. However, these outliers may be high-performing or low performing students.
that may need to be reflected in the teachers' effectiveness index. Further research is being conducted to gather information concerning the characteristics of these outliers.
References


