How Much is a Quality Education Worth?

The Implicit Market for Elementary Schools in Worcester, MA[1]

John C. Brown
Clark University

Saurav Katwal
Eastern Connecticut State University

December 1, 2015

ABSTRACT

This paper uses a two-stage approach to examine the effect of school value-added on the values of homes in Worcester, Massachusetts over the period 2006-2011. In the first stage, we estimate schools' value-added, i.e., schools' marginal effect on student academic achievement using student-level longitudinal data on test score performance. The effects of school performance are also measured for two groups: stable students do not change schools and mobile students who change schools or enter the system after third grade. We use empirical Bayes' estimates to account for measurement error in the value-added measures. In the second-stage, we identify the capitalization effect of the school value-added measures (estimated in the first-stage) on housing values within a traditional hedonic framework using boundary fixed effects to control for unobserved neighborhood heterogeneity. This paper finds some evidence for the capitalization of school value-added on housing values. In our sample, the school value-added on English scaled test scores is capitalized into the sales value of homes.

---

[1] This draft is for discussion purposes only. Please do not quote or cite without the permission of the authors. The address for correspondence is John Brown, Department of Economics, Clark University, Worcester, MA 01610 (jbrown@clarku.edu).We thank Jacqueline Geoghegan for sharing data on Worcester housing values. The Worcester Public Schools provided data and generous assistance with interpretation. The Mosakowski Institute for Public Enterprise and the Gurel Endowment offered generous research support. Katwal thanks the Clark University Department of Economics for summer research support.

Introduction

In comparison with other developed countries, schools in the United States rely disproportionately on local sources for the funding of education. Hanushek and Yilmaz (2012, Table 26.4) offer a summary of evidence on funding from the mid-2000s. Even high-poverty districts in suburban and rural areas, which can draw upon state and federal sources of aid, rely upon local resources for one-quarter of funding. Districts in high-poverty cities must cover up to 40 percent of their expenditures with local revenue sources. For all districts in the United States, data from 2008 suggest that a bit under one-half of school funding continued to come from local sources. In 15 states, the share was over one-half (see Massachusetts Budget and Policy Center (2010)).

The long-standing reliance on local sources for funding all local public services, the multiplicity of jurisdictions in most metropolitan areas and ongoing suburbanization led Tiebout (1956) to explore the consequences of a link between household mobility and local public goods provision. His conclusion was that mobility within metropolitan areas could serve as a relatively efficient way to allocate public goods that were well-matched with the tastes of local residents in terms of cost, variety and quality. Hoxby (2000) found, for example, that moving from a regime of no choice in schools to a regime of a plenitude of choices within a metropolitan area would raise school productivity and student achievement. The public finance literature, summarized a few years ago in Nechyba (2006), has explored the theoretical consequences of such a Tiebout sorting equilibrium under a variety of conditions. One key goal has been to identify reasons for why differences in quality across jurisdictions (or schools) could persist in the face of residential mobility. Key findings of the theoretical efforts focus on the existence of spatially hetereogenous housing stocks (from MacMansions in the suburbs to the ubiquitous triple-deckers of the center cities of Northeastern communities), which give rise to demands for location that vary by income. Substantial local contributions to schools add a second dimension. Finally, differences in non-financial inputs—peer effects or parental contributions—that are also likely to be correlated with income exacerbate the differences generated by heterogeneity in housing alone.[2] Nechyba (2006) notes that another strand of research focuses on transportation costs as a constraint on mobility that could generate similar results. Finally, he notes that theoretical exercises have underscored the potential adverse impacts on efficiency if the benefits of positive peer effects in wealthier school districts outweigh the costs to low-income districts. Hanushek (2014) offers an overview of distributional consequences, particularly arising from the differentials that arise in the development of cognitive skills.

Theoretical work and empirical studies provide strong evidence that sorting by income can induce the significant gaps in educational quality that are observed in the United States today. The continued strength of residential-based assignment systems across jurisdictions has served to reinforce the close link between educational quality (outcomes) and property values. Less clear is the extent to which such a link holds for intra-city property values given the typical gap between center-cities and the surrounding districts in terms of test score performance, income and racial and ethnic composition. This paper explores this question in the context of the second-largest school district in Massachusetts, the Worcester Public Schools. After a review of the main issues in the literature on measuring capitalization of school quality, it discusses alternative value-added measures of school quality that control for student characteristics. For the most part, these measures do not influence property values. However, it is noticeable that the

[2] See Nechyba (2006, pp. 1334-1335).

value-added measure for one group—mobile students—does have a significant impact on property values in the Worcester school district.

Measuring Capitalization of School Quality in Property Values

As Nguyen-Hoang and Yinger (2011) note in their comprehensive review of 50 of the most recent studies on this question, estimation of the extent to which school quality differences are capitalized in housing values faces a number of methodological challenges. The challenges arise from two sources: neighborhood heterogeneity that may be correlated with measured school quality and errors in measurement of school quality. For the purpose of argument, suppose that the value V_{it} of house i in school catchment zone or district k is related to a vector $\mathbf{X}_{ik}$ of housing characteristics, $\mathbf{Z}_k$ of neighborhood characteristics and q_{kt}, a measure of observed school quality in year t. Unmeasured attributes of the neighborhood l within catchment k are ζ_{kl} and actual school quality is $q_{kt} = \tilde{q}_{kt} + \omega_{kt,}$ or measurement error. A random disturbance $\varepsilon_{kit,,}$is distributed i.i.d. A standard hedonic regression such as (1) could be used to test for the capitalization of school quality in housing values:

$$(1)\ \ln (V_{it}) = \alpha + \beta(\tilde{q}_{kt} + \omega_{kt}) + \mathbf{X}_{ik}\boldsymbol{\varphi} + \mathbf{Z}_k\boldsymbol{\delta} + \zeta_{kl} + \varepsilon_{kit}$$

Even with a well-behaved random disturbance $\varepsilon_{kit,,}$ consistent estimation of the true effect (β) of school quality faces two obstacles that have been the subject of well over two decades of research attention. The first problem is one of unobserved neighborhood hetereogneity (ζ_{kl}). Unobservable neighborhood quality measures such as safe streets, well-tended park facilities, or neighbors who pick up their trash could be correlated with school quality. The standard solution in the literature to this problem has been to use boundary fixed effects (BFE) models first pioneered by Black (1999).

These models use a form of regression discontinuity design to capture unmeasured neighborhood effects by restricting the sample to properties on either side of a district or attendance zone boundary. The term ζ_{kl} is assumed to be constant across the boundary and estimated with a fixed effect for each boundary.

More recently, Bayer, Ferreira et al. (2007) and Kane, Riegg et al. (2006) have identified a second potential source of bias even within the context of a BFE design. Sorting among households could be positively associated with school quality or the unobservable characteristic ζ_{kl} could lead to the result that $\zeta_{kl} \neq \zeta_{jl}$ for the neighborhood on either side of the boundary between two school attendance zones k and j. In this case, the coefficient β is capturing both the unmeasured market valuation of ζ_{kl} and q_{l}. Hoang and Yinger (2011, p. 34) express skepticism for the approach adopted by Bayer and Ferreira, who include measures of income and other measureable household attributes that would be likely to be correlated with ζ_{kl}. Introduction of any of those variables (such as income) as a proxy for unmeasured neighborhood characteristics that may prompt sorting introduces a classic errors-in-variables problem within the context of hedonic regressions. Households simultaneously choose the entire bundle of attributes associated with a house, including school quality. Unmeasured influences on household demand not captured by income could also be correlated with demand for school quality, which introduces a potential bias in the estimation of β.

Finally, several issues associated with the appropriate measurement of $\tilde{q}_{kt}$ have received attention in the literature. One is whether the appropriate measures should be inputs or outputs. The first studies, starting with the pioneering piece by Oates (1969), focused on using spending or other measures of inputs. That approach has been generally abandoned in favor of using results on standardized test scores after the critique by Hanushek (1986) and others that there was little direction correlation between per-pupil expenditures and measureable outcomes. Most recent studies measure quality with outputs such as graduation, attendance or drop-out rates, student test scores, or school “grades” based on test scores. 39 of the 50 studies reviewed by Nguyen-Hoang and Yinger (2011) include measures of test scores.

In the context of the extensive literature on using test scores to measure teacher performance, Kane and Staiger (2002) argue that this measure of q_{kt} can be a noisy measure of $\tilde{q}_{kt}$ (the actual quality of the school). Along with random fluctuations in student performance, idiosyncratic events surrounding the taking of tests can impart errors that are common to a classroom of students or a grade. Small samples of test-takers for any classroom-grade-school unit of observation, particularly in elementary schools, could lead to relatively large measurement errors relative to $\tilde{q}_{kt}$. The measurement errors can lead to a downward bias to estimates of β provided these fluctuations are independent of ζ_{kl} and ε_{kit}. Kane and Staiger propose several adjustments discussed in more detail below that use empirical Bayes estimates to smooth out the impact of such fluctuations on measured differences across schools.

An important criticism of standard level measures of school performance focuses on the close correlation of test score performance with other variables such as parental income (the share of children receiving reduced or free school lunches), the share of students who speak English as a first language, and the share without learning disabilities. As performance on test scores became more closely associated with evaluation of school and teacher performance, the education community sought to identify other methods of measuring the quality of education provided in schools and classrooms.[3]Meyer (1997) Some studies examining the impact of school

[3] See Meyer (1997) for an early discussion. Tekwe, Carter et al. (2004) provide a review of models and Chetty, Friedman et al. (2011) offer estimates of longer-term impacts of value-added measures.

quality on property values have also argued that some measure of value-added—the contribution of the school and teachers to the student learning—rather than level measures should be used.

Most studies using measures that condition school performance on student or school inputs have utilized data available at a school or district level. Hayes and Taylor (1996) find that their value-added measure, which is based on data aggregated at the school level, is capitalized in housing values. Other earlier studies, including Brasington (1999), Downes and Zabel (2002), and Brasington and Haurin (2006), find that use of district-level value-added measures yields no evidence of capitalization. The value-added measures used in these studies are average gains measures, which are calculated by subtracting prior school or district level averages of test scores from current averages. These aggregate gains measures generally do not take individual student heterogeneity into account. Kane, Riegg et al. (2006) find no capitalization effect of their school value-added measure, which they estimate using individual student level data. Gibbons, Machin et al. (2013), however, find that in the United Kingdom, both the level and value-added measures of school quality are capitalized in housing values. Their measure of value-added is the difference between a student's test score at the end of key stage two (age 10-11) and her score at the end of stage one (age 7-8) aggregated at a school level. More recently, Imberman and Lovenheim (2013) estimate the capitalization effects of the release of value-added measures of school quality on house prices. Their study finds no evidence on the capitalization of two waves of new school and teacher value-added information released by Los Angeles Times (August 2010, and May 2011), or the school value-added information released by Los Angeles United School District (April 2011). This is a surprising result given the huge amount of media coverage received by these events. However, the authors find some suggestive evidence that the effect of value-added scores on house prices might be larger in lower-income neighborhoods. The school value-added estimates published by Los Angeles Times are based on Dynamic Ordinary Least Squares (DOLS) value-added models, which is similar to the models we use to estimate value-added of schools in Worcester. (see Guarino, Reckase, et. al. (2012)).

Value-added models of teacher and school quality

The discussion on the relevance of value-added measure in examining the capitalization of school quality on house prices is inconclusive. As explained above, some studies have found it to be significant and some have failed to find significant impacts. In more recent times, the use of student-level longitudinal data to estimate teacher and/or school quality has gained popularity among school districts. For example, Imberman and Lovenheim (2013) note that school districts in New York City, Houston, and Los Angeles provide both teacher and school value-added information to the general public. Massachusetts started to provide information on student "growth" by publishing a version of value-added measure starting in the 2008-2009 school year.[4] The use of value-added measures has become increasingly more important in the past few years due to their use in school and teacher accountability. Brasington and Haurin (2005) note that states are rapidly shifting from using school proficiency levels to value-added measures for accountability. This shift in school and teacher accountability has led to a rapidly growing literature on methods to estimate school and teacher value-added.

The studies of value-added models for teacher effects differ in their treatment of individual heterogeneity. On this basis, there are fundamentally two kinds of value-added models: the gains model and the lagged-score models. The gains models use gains in scores (year-to-year change in student score) as the dependent variable along with student fixed effects

[4] Information on Massachusetts' value-added growth measure can be found at http://www.doe.mass.edu/mcas/growth/. An important distinction between this measure and the measure based upon individual student data developed in this study is that it compares an individual student's year-to-year performance against the population of students who performed similarly in the index year.

to address individual student heterogeneity.[5] The lagged-score models use the year's score as the dependent variable and account for student heterogeneity by including a lag of student test score.[6]

The main debate in the teacher value-added literature is whether or not the estimates of teacher quality yielded by value-added models are biased or unbiased. Kane and Staiger (2008) found that teacher effects estimated using a value-added model that included prior student test scores were unbiased predictors of teacher effects in an experiment where teachers were randomly assigned to students, in Los Angeles Unified School District (LAUSD). Comparing different value-added models of teacher effects in their simulation exercise, Guarino, Reckase et al. (2012) find that the lagged score models are the most accurate in predicting teacher effects. As the lagged score model controls for past achievement, it is effective in dealing with bias in teacher effects due to students being sorted into classrooms or teachers based on past achievement. Rothstein (2010) conducts a falsification test of teachers' value-added and reports that the standard value-added models including the lagged-score model show a significant effect of teacher quality on past student achievement, thereby rejecting a causal interpretation of estimates of teacher effects yielded by value-added models. The main implication of Rothstein's result is that teachers might be assigned to students based on unobserved and time-variant student attributes which neither lagged scores nor student fixed effects account for. Several other studies, however, suggest that the value-added models can yield reliable measures of teacher effects under certain conditions. For example, Koedel and Betts (2011), while accepting Rothstein's general critique, suggest that value-added models based on multiple years of data that use student fixed effect, and exclude novice teachers (teachers in their first two years) can

[5] See Hanushek, Kain, and Rivkin (2005), Jacob and Lefgren (2008), and Boyd et al. (2007).
[6] See Aaronson et al. (2007), Kane, Rockoff, and Staiger (2008), Jacob and Lefgren (2008), Boyd et al. (2007).

significantly mitigate the problems associated with non-random sorting of students and teachers. Using their simulation exercises, Goldhaber and Chaplin (2012) question the validity of Rothstein's falsification test itself, as they report that the Rothstein falsification test rejects value-added models even in the absence of bias in estimated teacher effects. Chetty, Friedman et al. (2011) find that there is minimal or no bias in teacher effects due to sorting on teacher and students' observable attributes, and no bias at all due to sorting on unobservable attributes.

This study focuses on housing market valuation of school quality under the assumption that new home buyers more likely are choosing a school rather than an individual teacher. Sanders (2000) notes that value-added estimates of *school* quality are less susceptible to bias than estimates of teacher effects Sanders (2000). In addition, the estimation of teacher effects requires higher level of data quality and methodological sophistication because it suffers from smaller sample size i.e. smaller number of students per class, which makes it more vulnerable to sampling error than an estimate of school effects.

Apart from their treatment of individual student heterogeneity, value-added models also differ in the way that they specify teacher effects. In particular, teacher effects can either be modeled as fixed effects or random effects. Buddin and Zamarro (2009) specify teacher effects as fixed effects, whereas other studies including Kane and Staiger (2008), McCaffrey, Lockwood et al. (2003), and Chetty, Friedman et al. (2011) model teacher effects as random effects. Raudenbush and Bryk (2002) point out that the random effect models might be preferable (to fixed effects models) as they can produce the best linear unbiased predictors (BLUP) or empirical Bayes estimators. The empirical Bayes estimates of teacher or school effects can be obtained by using Hierarchical Linear Models (HLM), which yield consistent and efficient maximum likelihood estimates of the parameters of interest in models with nested random effects. However, most studies including Kane and Staiger (2008) and Chetty, Friedman et al. (2011) use ordinary least squares (OLS) and perform Empirical Bayes adjustment to account for some of the measurement error issues in a residual based model. Buddin (2010) applies similar Empirical Bayes adjustment to his estimates of teacher and school fixed effects. The Empirical Bayes estimates are calculated by multiplying the estimated school effects by its reliability factor. The estimated reliability factor for each school effect is the ratio of adjusted variance of school effects across schools and the sum of the adjusted variance and squared standard error of the school effect. The Empirical Bayes adjustment can be performed irrespective of whether teacher or school effects are estimated as fixed or random effects. It should be noted, however, that these adjusted estimates are not the "true" Empirical Bayes estimates. The true Empirical Bayes estimates are based on the second Hierarchical Linear Model, where the school effects are estimated as random effects.

Tekwe, Carter et al. (2004) provide a comparison of the simple fixed effects (school effect specified as fixed effects in an OLS framework), and Hierarchical Linear Models (HLM). The authors find that the estimates of school effects obtained from these models are highly correlated with each other. The authors indicate that the choice between these two models depends on the answer to the following question: Should schools be held accountable for the effects of their socio-demographic composition on the value-added of its students? If the answer to this question is yes, then the HLM is preferred. However, the schools might be partially responsible for these socio-demographic effects or "peer effects", which would make the choice between these two models difficult.

The Empirical Model

The primary goal of this paper is to test whether housing markets also capitalize value-added measures of school quality. The approach uses student-level longitudinal data and focuses on testing whether quality measures estimated for elementary schools are capitalized in the sales price of housing. Therefore, the first part of this section discusses the empirical model of school quality capitalization and the second section discusses alternative estimation strategies for value-added models of school quality.

1. Hedonic Model of School Quality Capitalization

The empirical model used in this study is based on the framework first articulated by Rosen (1974) that an explicit market for bundled goods such as housing is made up of a number of implicit markets, for example, school quality, aversion to crime, and so on. As Nguyen-Hoang and Yinger (2011) and others have emphasized an important implication of the Rosen framework first developed in his 1974 paper: the hedonic estimation method used to recover information about capitalization provides only information about the implicit marginal price of an attribute of the bundled good and reveals nothing about the underlying demands. Thus the hedonic price function found in equation (1) can be used to test for capitalization. In this application, the vector X_{ik} includes such characteristics of house i as the square footage, the number of bathrooms and bedrooms in the house, the age of the house, and the presence of amenities such as a porch, outbuilding, attic and air-conditioner. The house attributes and amenities are expected to be positively associated with house values with the exception of the age of house. Z_k is a vector of neighborhood characteristics. As in Sandra E. Black (1999), we address the unobserved neighborhood heterogeneity by modifying the model:

$$(2)\ \ln(P_{it}) = \alpha + \beta(\tilde{q}_{kt} + \omega_{kt}) + \mathbf{X}_{ik}\boldsymbol{\varphi} + \mathbf{Z}_k\boldsymbol{\delta} + \mathbf{K}_b\theta + \varepsilon_{kit}$$

where K_b is a vector of b boundary dummies (b>k). In the analysis that follows, we focus on samples that use a buffer of 0.10 miles around the attendance zone boundaries. Figure 1 provides a map of the elementary school attendance zones and property sales within 0.1 mile of zone boundaries. Houses within 0.1 mile of attendance zone boundaries are about one to two blocks from the boundary. This methodology is conceptually equivalent to calculating differences in mean house prices on opposite sides of attendance zone boundaries and relating this to differences in test scores, where the boundary dummies account for any unobserved housing and/or neighborhood characteristics shared by houses on either side of the boundary. In all of the model specifications, sale year dummy variables control for year-to-year change in housing prices. Initial specifications included other location-specific attributes such as proximity to a four-lane highway, the distance to the city center, and distance to the elementary school. None of these variables had a statistically significant coefficient so they are excluded from the results presented here. We also found the indicator variables for housing types to be individually and jointly statistically-insignificant, and the same was true for middle school value-added estimates; hence, we exclude these variables in the final set of model specifications.[7] Since we have multiple housing sales transactions in each school attendance zone area at different points in time, all model specifications are corrected for clustering at the school attendance zone level.

2. Value-Added Models of School Quality

Standardized test scores used in most of the capitalization literature are not adequate as measures of true school quality because they are aggregate measures that capture a number of different influences: family background, student ability and peer effects as well as the marginal

[7] The inclusion of middle school value-added estimates have no effect on the effect-size of elementary school value-added measures.

effect of school or "school quality". Yearly average school-level test scores are highly correlated with students' socioeconomic status such as poverty level, minority status, and language proficiency level. Table 3 provides information on the quartile distribution for elementary schools of an index used in Massachusetts to summarize standardized test results of students across schools, the Composite Performance Index (CPI).[8] Notably, 86% of students in schools that are in the lowest quartile are eligible for free or reduced price lunch compared to 38% in highest quartile schools. Schools in the top quartile of CPI also have a significantly lower proportion of students with limited English proficiency (8%) and minority students (20% Hispanic and 10% African American) than the bottom quartile schools, where a quarter of the students have limited English proficiency, and 44% and 13% of students are Hispanics and African Americans respectively. This is the key motivation for using value added method to get at the marginal effect or the true effectiveness of schools.

First developed in the education assessment literature, all of the value-added models we use are based on a general lagged test score (dynamic ordinary least squares) model as given by the equation below

$$(3) \quad Y_{it} = \lambda Y_{i,t-1} + \mathbf{S}_{kt}\mu + X_{it}\gamma + e_{it},$$

where Y_{it} and Y_{it-1} are current and lagged student test scores, $\mathbf{S}_{kt}$ is a vector of school-year dummy variables, $\mathbf{X}_{it}$ is a vector of time variant classroom and individual characteristic such as free-lunch eligibility and limited proficiency in English and e_{it} is an idiosyncratic error.
We estimate three value-added models of school effects and ignore classroom-specific effects; we assume that potential purchasers of homes are not choosing a particular teacher, but that they

[8] The CPI is a weighted average of the proportion of students falling in each of four performance categories ranging from failure to proficient.

are concerned about overall school quality. Each model varies in terms of the ways it treats the school-year effect and the error term. The first model (VAM 1) specifies school effects as fixed effects and individual student effects as random effects (FGLS). The second model (VAM 2) specifies school effects as fixed effects, but assumes that the individual student error terms follow an AR 1 process, and uses maximum likelihood estimation (MLE). The third model (VAM 3) is a Hierarchical Linear Model (HLM), which specifies school effects as random effects and yields the "true" empirical Bayes estimates. The vector of year and school effects ($\boldsymbol{\mu}$) estimated using some variation of the above model is then used to provide the value-added measures of school quality in the hedonic price equation.

Since one of the independent variables in the model is the lag of a student's test scores, estimates of school value-added are potentially biased due to measurement error in student test scores. A standard practice in the literature to mitigate this bias is to perform Empirical Bayes adjustment to account for some of the measurement error (see (Koedel et al., 2012)). The Empirical Bayes adjusted estimates are calculated by multiplying an estimated school-year effect by its reliability factor. The estimated reliability factor for each school effect is the ratio of adjusted variance of school effects across schools and the sum of the adjusted variance and squared standard error of the school effect. We apply the empirical Bayes adjustment to the estimates of school value-added obtained using models 1 and 2. Although, VAM 1 and VAM 2 are structurally similar, there is a key difference between the two; VAM 2 uses Maximum Likelihood Estimation, whereas VAM 1 is modeled as Feasible Generalized Least Squares (FGLS) with student effects as random effects.

If the school effects are random effects in the true model, the OLS estimates are consistent, but inefficient. An alternative to OLS is a Hierarchical Linear Model (HLM), which is optimized to estimate nested random effects as it uses Maximum Likelihood Estimation (MLE) and thus provides efficient estimates of teacher and/or school effects (see Raudenbush and Bryk, 2002). Our third school value-added model (VAM 3) is an HLM; it specifies school effects as random effects and yields the Estimated Best Linear Unbiased Predictor (EBLUP) or the "true" empirical Bayes. Adcock and Phillips (1997) note that HLM allows the examination of associations among multi-level, nested data such as students within schools by estimating simultaneous linear equations at the student level within schools and the school level between schools.

The HLM framework draws a clear conceptual distinction between student-level and school-level effects. The student-level model estimates effects at the student level (Level 1) within each school, and the school-level model (Level 2) explains the student level effects in terms of school effects. The result is that school effects are estimated as random effects. Consider the two-level system of equations below:

Level 1:

(4) $Y_{ikt} = \pi_{0kt} + \pi_{1k}Y_{i,t-1} + \pi_{2k}\, g_{ikt} + \pi_{3k}\, l_{ikt} + \pi_{4k}\, p_{ikt} + e_{ikt}$

Level 2:

(5) $\pi_{0kt} = \gamma_0 + \zeta_{0kt}$

(6) $\pi_{1k} = \gamma_1$, $\pi_{2k} = \gamma_2$, $\pi_{3k} = \gamma_3$ and $\pi_{4k} = \gamma_4$

Where,

Y_{ikt} – Test Score of Student "i" in school "k" in year "t".
π_{0kt} – School-year specific random intercept.
$Y_{i,t-1}$ – Test Score of student i in the previous year.
g_{ikt} – Grade year indicator variable.
l_{ikt} – English language proficiency status of student in the current year.
p_{ikt} – Free/Reduced Lunch eligibility status of student in current year.
γ_0 – Predicted grand mean of achievement for all schools over all years of data.

ζ_{0kt} – Effect of school k on the expected school achievement in year t.

The above two-level HLM model is a random intercepts model. The intercept π_{0kt} is the only random component (besides the idiosyncratic error term). Level 2 specifies a separate equation for π_{0kt}. It should be noted that the other level-1 independent variables could also be specified as random; however, we assume that all of the other coefficients in level 1 are fixed throughout the system, as shown by equation (6). This assumption implies that the effect of individual student attributes such as lagged test scores do not vary across schools.

A number of variables associated with school context, such as socio-demographic composition of schools, and a number of other variables associated with school practices, such as teacher quality or after-school programs or recreational programs could be affecting students' academic growth (see Adcock and Phillips (1997)). Some studies include school or classroom level variables such as the school-level average of student socio-economic status, and teachers' experience and teachers' use of alternative assessments to determine particular factors that contribute to the value-added by schools (see Alkharusi (2011)). As noted above, we exclude school-level attributes in level 2, the school context variables, such as the percentage of poor or minority students, because our primary focus is to estimate an aggregate measure of school value-added.

The justification for using HLM in estimating school effects comes from the high likelihood of the correlation between errors on the measurement of student achievement within each school, which would violate the standard linear regression assumption that e_{it} is i.i.d. HLM's two-level estimation of random effects model accounts for the non-independence in errors. HLM's ability to partition the random component of the model into within and between school variability is not a feature that is shared by standard approaches such as ANCOVA. In addition, the random effects estimates of school value-added are the actual Empirical Bayes estimates of school effects, which are more efficient than estimates derived from OLS. The empirical Bayes estimate of the school effect, ζ^*_{0kt} is based on the following empirical Bayes estimate of random school intercept,

(7) $\pi^*_{0kt} = r_k [Y_{kt} - \pi_{1k} Y_{i,t-1} - \pi_{2k} G_{ikt} - \pi_{3k} E_{ikt} - \pi_{4k} P_{ikt}] + (1 - r_k) [\gamma_0]$

Alexander, Entwisle et al. (1996) note that the empirical Bayes estimate of a school value-added is a weighted average of the level 1 (within-school) estimate and the predicted grand mean of achievement for all schools. In equation (7), as the reliability factor (r_k) approaches 1, the estimate from the within-school model (level 1) provides the estimates of school effects (Alexander, Entwisle et al. (1996)). If the reliability factor is negligible (close to zero), then the estimate of the school effects is the predicted grand mean of achievement of all schools. The model-based or empirical Bayes estimates are penalized or "shrunk" if they are estimated with low precision (a large standard error). Again, the most relevant example is the case where the school-effects estimates are noisy due to a small number of students. Essentially, the empirical Bayes estimation method pulls the estimates towards the mean, or, it "borrows strength" from the mean, which in this case is a grand mean.

Data Description

This paper focuses on elementary schools in the Worcester Public Schools District. The Worcester public school system currently includes 31 elementary schools and the admissions to these schools are primarily based on the residential locations of the students.[9] Hence, the city is divided into 31 elementary school attendance zone areas. Intra-district school choice is available,

[9] Given the focus of this paper on the value-added measures of school quality, the analysis covers the period 2006 through 2011. The Massachusetts Department of Education began administering the Massachusetts Comprehensive Assessment System (MCAS) tests to four elementary grades (three through six) in 2006.

but it is apparent that the students/families that actually practice this choice represent a small fraction of the entire population of the elementary public school students in Worcester. We use the geographical information systems software ArcGIS to create the attendance zone map. Data on the website of the parent information center of Worcester Public Schools matches all address ranges in the city to a school attendance zone.[10]

Our housing data include housing sales data from the Warren Group, which includes the housing transactions for the (lagged) period for which test score data are available. We focus on the sales of three types of houses; single or 1-family houses, 2-family houses and 3-family houses (many of which are triple-deckers. Restricting the transactions to the relevant time period, housing structures and 0.1 mile buffers, we are left with a bit over 3,300 sales transactions. Two-thirds of the sales are of single-family residences.[11] Three-family residences make up the second largest group with 21%, and the rest of them are two-family residences (13%). The real house prices are in 2005 dollars, which are calculated using consumer price index data from Bureau of Labor statistics.[12] Table 1 provides a summary of the housing characteristics that were included in the study.

The administrative records of the Worcester Public Schools (WPS) provide the student-level data used in the estimation of school value-added measures. The data include the individual MCAS scores and other student information such as the grade, class, school, and student demographics. A randomly generated identifier was assigned to each student, which allowed creation of a longitudinal dataset on math and English language test scores for students taking

[10] The URL for the website is http://pic.worcesterschools.org/.
[11] We dropped from consideration transactions less than or equal to $10,000.
[12] The URL for the website is http://www.bls.gov/data/#prices.

the MCAS tests in grades 4 through 8 and other attributes for elementary students (grades 2 through 6) for the period 1998-2011.[13] Only anonymized data were used in the analysis.

Table 1 presents the descriptive statistics for the housing price data. Columns 1, 2, and 3 present the number of observations, mean and standard deviation for the key variables and columns 3, 4, and 5 do the same for our boundary sample (0.10 miles within attendance zone boundary). Table 1 shows that the mean house prices in the full sample are identical to that of the boundary sample. The mean test score measures, however, are slightly lower in the boundary sample. Houses in the boundary sample tend to be smaller and older on average than the full sample.

Table 2 shows the descriptive statistics for the variables used in the value-added models of school quality. The average English scaled score is slightly larger than the average Math scaled score; 236 and 232 respectively. In our sample of students, 63 percent are eligible for free or reduced lunch and 16 percent have limited proficiency in English.

Each housing transaction is paired with the latest school value-added measure. We keep the assignment of school value-added measures to sales transactions consistent with the public release cycle of MCAS test scores. Given that the MCAS test scores are reported in the September of each year (the tests are administered in the spring), we assign the most recent value-added measure to all the transactions that take place between the October of the same year and September of next year. For example, we assign 2010 value-added measure, which uses MCAS test scores from 2009 and 2010, to all the house sales transactions that take place

[13] Proper steps have been taken to ensure protection of human subjects (Note: Prof. Brown and provide more information on this).

between the October of 2010 and September of 2011. MCAS scores for 2011 were used for the final period of housing sales, which lasted into the end of 2011.

Results from the Value-Added models

Table 3 presents results from the value-added models of school effects. The school effects are jointly significant in all of the models, and the standard deviation of the empirical Bayes' adjusted school effects range between 0.10 and 0.16 across different models. All models control for lagged test scores and other student characteristics; they include dummy variables for grades. As expected, the lagged scores and free/reduced lunch eligibility are positively and negatively associated with test scores, respectively. The effect of past achievement on current test scores ranges from 0.64 to 0.86 across models. Not surprisingly, limited English proficiency has a negative effect on English test scores. The effect of limited English proficiency on the math score varies with each model; the coefficients are negative and significant, positive and significant, and positive and insignificant, in the first, second, and third models, respectively.

The estimates of the school effects from the three different value-added models (VAM1, VAM 2, and VAM3) are highly correlated with each other.[14] For example, Table 4 shows that the correlation between average school effects range between 0.94 and 0.98.

Table 4 also shows that the correlation between the value-added measures and level measures of test scores is quite low, which potentially makes value-added measures less correlated with students' socio demographic attributes, and hence better measures of schools' marginal effect on students. Table 4 shows that the correlations between different value-added

[14] Note that the table presents the results for empirical Bayes adjusted versions of VAM1 and VAM2.

and the composite measure of levels of test score performance (the CPI measures) range between 0.25 and 0.41. Presumably the difference between value-added measures and the CPI measures is the strong influence of student socio-economic characteristics on test score results.

One way to explore whether this is indeed the case is to look at the extent to which measures of value-added identify a different group of high-performing schools compared with the CPI measures. We calculated the residuals from a simple regression of the value-added measures for math and English on the respective CPI for elementary school-year pairs. As indicated by the correlation coefficients presented in Table 4, the R^2 for these regressions were in the range of 0.10 to 0.18.[15] Figure 2 shows a graph of the schools that were not low-performing (at most one year with a value-added measure in the first quartile for both areas) and high-performing in either English or math. Figure 3 shows a similar graph of schools that were low-performing in either English or math and rarely in the top quartile in at most twice in either test category. Given that the median CPI for Worcester schools is about 67 during this period on both tests and the top quartile is at about 72-75, it is clear that a small number of schools show exceptional performance beyond what is predicted by the CPI of that school (numbers 10, 6, and 27). In addition, we can identify a few schools (12, 20 and 21) that appear to underperform when compared with what would be expected based upon average MCAS test scores.

One potential weakness of value-added measure, which it shares with the level measures of test scores, is that it can be inter-temporally unstable. Sampling error due to small class sizes could cause the value-added measures to be unstable. Empirical Bayes adjustment of the value-added measures can potentially account for some of the measurement error (Buddin, 2010). Table 5 compares different value-added measures in terms of their inter-temporal instability. For

[15] The regressions are EnglisVA=-4.17+0.058*EngCPI (with an R^2 of 0.10) and MathVA=-5.75+0.085*MathCPI (with an R^2 of 0.18).

each year between 2006 and 2011, each school was ranked on the basis of average test score level, and various value-added measures. Table 5 shows the proportion of schools that ranked in the top quartile for different number of years for various measures of school quality. If the schools were perfectly stable, or, if the schools' rankings could be determined perfectly then, 25% of schools would rank in the top quartile in all 7 years and 75% of schools would never rank in the top quartile. It is apparent from the rankings that the level measures as well as all the value-added measures are quite unstable across the years. For example, when ranked on level measures, 50 percent of schools didn't rank in top quartile in any of the years, and only about 13 percent of schools ranked in the top quartile in at least 4 years. Similarly, when ranked on the model 1 average value-added measure, 49 percent of schools never ranked in the top quartile, and 20 percent ranked in the top quartile in at least 4 years. Empirical Bayes adjusted measures from model 2, and estimates from model 3, seem to identify a slightly larger proportion of schools as consistently high value-adding schools, however, the difference is quite small.

Results from Hedonic Estimation

Table 6 reports results from hedonic regression specifications that use Empirical Bayes' adjusted school value-added measures estimated using Maximum Likelihood Estimation (MLE).[16] The school value-added measures in the first 3 columns are based on all students; those in the middle 3 columns (columns 4 through 6) are based on models that include only those students who don't change elementary schools (stayers) and the last three columns are based on

[16] We don't report results from hedonic regressions with school effects estimated from FGLS, because the results are quite similar, the two measures are highly correlated and the value-added estimates from MLE are more efficient; however, the results from FGLS specifications are available upon request.

models that include the transient students who switch between elementary schools or are new to the Worcester Public School System (movers).

Table 6 shows that the coefficients on all of the housing attributes have the expected signs. The coefficients on the number of bathrooms, bedrooms, lot size, and dummy variable indicating the presence of air-conditioner are all positively associated with house values and are measured with small errors. The results consistently show across different specifications that the age of a house is viewed unfavorably by the home-buyers as it has a negative and significant effect on house values.[17] More importantly, Table 6 shows that the value-added measures of school quality based on English test scores are significantly associated with house prices, and there seems to be no significant difference between the capitalizations of value-added measures estimated using the stable students or stayers, and mobile students (movers).

As discussed above and exhibited in table 5, the school value-added measures are highly unstable going from one year to the next. To account for instability in value-added measures due to measurement error, the results in Table 6 use the empirical Bayes' adjustment of the school value-added estimates. For estimates of Value-Added that do not distinguish between stayers and moves, the capitalization effect for the English value-added measure is significant at better than 10 percent for a one-tail test. A 1 standard deviation increase in value-added measure estimated using English scaled test scores of all students (movers and stayers) is associated with 1.9% increase in house prices (see column 1). The coefficients on the math and average value-added scores are very small and are not statistically significant. There isn't a substantial difference in

[17] This and subsequent specifications of the hedonic relationship does not control for spatial variables such as distance to city center, distance to the school or proximity to a four-lane highway. Specifications that included these spatially-specific variables did not improve the fit, and t- and F-tests found them to be individually and jointly insignificant.

magnitude between the coefficient on the English value-added measure for stayers (see column 4) and the coefficient for the same measure based on transient students or movers (see column 7). A one standard deviation increase in the school value-added measure based on scaled English scores of stable students and one standard deviation increase in the same measure based on transient students are associated with 1.5% and 1.46% increase in house prices respectively.

Table 7 presents the results of regressions that use an alternative measure of value-added estimated using a hierarchical linear model (HLM). These results generally support the earlier finding, although significance levels for measures calculated using results for all students have slipped slightly below ten percent and the coefficients are a bit lower. For all students, a one standard deviation increase in the value-added of a school is associated with 1.7 % increase in house prices. It should be noted that the results are quite similar when we exclude one of the schools in our sample, which is a magnet school and thus not strictly neighborhood a school.[18]

We further examine the capitalization of school value-added by estimating an alternative specification where we try to estimate the treatment effect of being on a high value-added school as opposed to a low value-added school. Table 8 (column 1) reports results on the treatment effect of being on the better side (side with higher school value-added) of the catchment boundary. The specifications in Table 8 are different from the earlier specifications in that they replace the estimated school value-added measures with a dummy variable which takes a value of 1 if the property is in higher value-added side and 0 otherwise. Column 1 of Table 8 shows a 2.4 % premium associated with being on the side of the boundary that has the school with higher value-added relative to the school across the boundary. However, the coefficient is not

[18] Worcester Arts Magnet school is one of the three magnet schools in the city. However, unlike the other two magnet schools, this school has been assigned an elementary attendance catchment area.

statistically significant. As in the earlier specifications the model includes boundary fixed effects, and the sample is restricted to houses within 0.10 miles of school attendance zone boundaries.

We find some evidence on the capitalization of school value-added on house prices. To check for the robustness of this finding, we perform a falsification test where we compare houses within the same school catchment. The concern is that the coefficients might be picking up the effect of progressive change in neighborhoods instead of differences in school quality as measured by school value-added. Using a method similar to the one used in Black (1999), we test this hypothesis by creating artificial boundaries within each elementary school catchment area.[19] Column 2 of table 8 shows the results from this falsification test. If the school value-added measures were really capturing the progressive change in neighborhoods, then the coefficient on the dummy variable for being on the high-side of the artificial boundary would also be positive. As expected, however, the coefficient on the artificial high side dummy variable is negative. This result makes us more confident that - to the extent that it is capitalized – the coefficients on the school value-added measures reflect the association between school quality and house prices rather than association between progression or change in neighborhood quality and housing values.

As a further check on the robustness of the capitalization of school value-added on students' performance on English MCAS tests, we extend our analysis by replacing our estimated value-added measures with an alternative measure of school value-added. Since 2008,

[19] Houses that are originally associated with the catchment area with the better school and that are within 0.10 miles of the actual boundary are reassigned as being on the "worse-side" of the artificial boundary, and the houses associated with the catchment area with better school, and which are between 0.10 and 0.20 miles of the actual boundary are reassigned as being on the "high/better-side" of the artificial boundary. The reassignment is exactly opposite for the houses that are actually associated with worse schools: those within 0.10 miles of the actual boundary are reassigned as being on the "better-side" and those between 0.10 and 0.20 miles are reassigned as being on the "worse-side" of the artificial boundary.

Massachusetts Department of Education (DOE) started releasing its own measure of value-added for each school - the median student growth percentile (SGP) - every September along with the results on the standardized MCAS test results.

Table 9 presents the results from the hedonic regressions which include SGP instead of or, in addition to our estimated school value-added measure. The results support our initial finding that the school value-added in English MCAS scores – as measured by our estimates from school value-added models – are capitalized into house values. Specifications in columns 3-5 control for neighborhood heterogeneity with boundary fixed effects, and dummy variables for census block groups and/or police statistical areas (PSA). When both of the school value-added measures – school value-added estimated from school value-added models and the Mass DOE published SGP – are used in the same regression, only the estimated value-added measure has a statistically significant effect on house values. We find that a one standard deviation increase in the school value-added measure based on scaled English scores is associated with 1.7% increase in house values. The coefficient on the Mass DOE published SGP is not statistically significant in these specifications (columns 3 through 5).

These results are quite remarkable, as they point to a few interesting aspects of the interaction between the housing market and public elementary schools in Worcester. The results strongly suggest that home-buyers and the Worcester community at large are aware of the elementary schools' ability in raising MCAS test scores; the coefficients on the estimated school value-added on MCAS English tests are statistically significant and are quite robust. The results also suggest that the community is largely unaware of Mass DOE's publication of SGP measures. Alternatively, it might be the case that the community is aware of the SGP measure, but this new information does not alter their perception on the ability of elementary schools' in raising students' academic performance. It should be noted that the estimated school value-added and the SGP measures are positively correlated, but the correlation is quite low (0.20).

Conclusions

Studies have consistently found that the level measures of test scores (school output or school quality) are capitalized on property values. The magnitude of capitalization, however, varies across studies. For example, in a across district study, Brasington and Haurin (2006) find that a one standard deviation increase in district level test score average is associated with around 7.1% increase in house prices. Studies examining the within district – across attendance zone capitalization find somewhat smaller capitalization effects. Black (1999) and Bayer et al. (2007), for example, find that a one standard deviation increase in school-level average of test score levels are associated with 2.1% and 1.8% increase in house prices, respectively. Katwal (2014) finds that a one standard deviation increase in level test score (average) measures is associated with about 4.5% increase in house prices.[20] Most studies that use value-added measures of school quality, such as Brasington (1999), Brasington and Haurin (2006), Kane et al. (2006), and Imberman and Lovenheim (2013) find no evidence for the capitalization of value-added measures of test scores on house values. Using a student-based measure of value-added with a panel of high quality test score data from the Worcester Public Schools, we find some evidence on the capitalization of school value-added. In our sample, we find that school value-added on English test scores are capitalized on house prices– a 1 standard deviation increase in the school value-added is associated with about two percent increase in sales prices.

[20] Katwal (2014) is a working paper by the author, and the results from the paper are available upon request.

These results suggest that housing markets in a medium-sized center city with a diverse student population do capitalize refined measures of educational quality. Further research on this issue will examine in greater detail whether value-added measures based on different subgroup of students have different capitalization effects and the extent to which student mobility is responsive to the diverse range of educational opportunity available in the city. Another natural extension of this analysis would be to further examine in greater detail, the capitalization of the school value-added measure – student growth percentile (SGP) - released by Massachusetts Department of Education (DOE).

References

Adcock, E. P. and G. W. Phillips (1997). "Measuring school effects with hierarchical linear modeling: Data handling and modeling issues." *Multiple Linear Regression Viewpoints* **24**(1): 1-10.

Alexander, K. L., D. R. Entwisle, et al. (1996). "Children in Motion: School Transfers and Elementary School Performance." *The Journal of Educational Research* **90**(1): 3-12.

Alkharusi, H. (2011). "Hierarchical linear models: Applications in educational assessment research." *Educational Research Journal* **26**(1): 41.

Bayer, P., F. Ferreira, et al. (2007). "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* **115**(4): 588-638.

Black, S. E. (1999). "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* **114**(2): 577-599.

Brasington, D. and D. Haurin (2005). Capitalization of Parent, School, and Peer Group Components of School Quality into House Price, Department of Economics, Louisiana State University, Departmental Working Papers.

Brasington, D. and D. R. Haurin (2006). "Educational Outcomes and House Values: A Test of the value added Approach*." *Journal of Regional Science* **46**(2): 245-268.

Brasington, D. M. (1999). "Which Measures of School Quality Does the Housing Market Value?" *Journal of Real Estate Research* **18**(3): 395-413.

Buddin, R. (2010). "How effective are Los Angeles elementary teachers and schools?".

Buddin, R. and G. Zamarro (2009). "Teacher qualifications and student achievement in urban elementary schools." *Journal of Urban Economics* **66**(2): 103-115.

Chetty, R., J. N. Friedman, et al. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood, National Bureau of Economic Research.

Downes, T. A. and J. E. Zabel (2002). "The Impact of School Characteristics on House Prices: Chicago 1987-1991." *Journal of Urban Economics* **52**(1): 1-25.

Gibbons, S., S. Machin, et al. (2013). "Valuing School Quality Using Boundary Discontinuities." *Journal of Urban Economics* **75**: 15-28.

Goldhaber, D. and D. Chaplin (2012). "Assessing the 'Rothstein Falsification Test': Does It Really Show Teacher Value-Added Models Are Biased?" *Center for Education Data & Research Working Paper*.

Guarino, C., M. D. Reckase, et al. (2012). "Can value-added measures of teacher education performance be trusted." *East Lansing, MI: Education Policy Center at Michigan State University*.

Hanushek, E. (2014). Is Location Fate? Distributional Aspects of Schooling. *Education, land, and location*. G. K. Ingram. Cambridge, Massachusetts, Lincoln Institute of Land Policy**:** 25-61.

Hanushek, E. A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* **24**(3): 1141-1177.

Hanushek, E. A. and K. Yilmaz (2012). Urban Education: Location and Opportunity in the United States. *The Oxford Handbook of Urban Economics and Planning*. N. Brooks, K. Donaghy and G.-J. Knaap, Oxford and New York: Oxford University Press**:** 583-615.

Hayes, K. J. and L. L. Taylor (1996). "Neighborhood School Characteristics: What Signals Quality to Homebuyers." *Federal Reserve Bank of Dallas Economic Review*: 2-9.

Hoxby, C. M. (2000). "Does Competition among Public Schools Benefit Students and Taxpayers?" *The American Economic Review* **90**(5): 1209-1238.

Imberman, S. A. and M. F. Lovenheim (2013). Does the Market Value Value-Added? Evidence from Housing Prices After a Public Release of School and Teacher Value-Added, National Bureau of Economic Research, Inc, NBER Working Papers: 19157.

Kane, T. J., S. K. Riegg, et al. (2006). "School Quality, Neighborhoods, and Housing Prices." *American Law and Economics Review* **8**(2): 183-212.

Kane, T. J. and D. Staiger (2002). "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." *Brookings Papers on Education Policy* **5**: 235–283.

Kane, T. J. and D. O. Staiger (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation, National Bureau of Economic Research, Inc, NBER Working Papers: 14607.

Katwal, S. (2014). "Linkages between Housing Market and an Implicit Market for Education." Working Paper.

Koedel, C. and J. R. Betts (2011). "Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique." *Education* **6**(1): 18-42.

Koedel, C., R. Leatherman and E. Parsons (2012). "Test measurement error and inference from. value-added models." The BE Journal of Economic Analysis & Policy **12**(1).

Massachusetts Budget and Policy Center (2010). Public School Funding in Massachusetts: Where We Are, What Has Changed, and How We Compare to Other States. *Policy Briefs*. Boston, MA, Massachusetts Budget and Policy Center.

McCaffrey, D. F., J. Lockwood, et al. (2003). *Evaluating Value-Added Models for Teacher Accountability. Monograph*, ERIC.

Meyer, R. H. (1997). "Value-added indicators of school performance: A primer." *Economics of Education Review* **16**(3): 283-301.

Nechyba, T. J. (2006). Chapter 22 Income and Peer Quality Sorting in Public and Private Schools. *Handbook of the Economics of Education*. E. Hanushek and F. Welch, Elsevier. **Volume 2:** 1327-1368.

Nguyen-Hoang, P. and J. Yinger (2011). "The Capitalization of School Quality into House Values: A Review." *Journal of Housing Economics* **20**(1): 30-48.

Oates, W. E. (1969). "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." *Journal of Political Economy* **77**(6): 957-971.

Raudenbush, S. W. and A. S. Bryk (2002). *Hierarchical linear models: Applications and data analysis methods*, Sage.

Rosen, S. (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* **82**(1): 34-55.

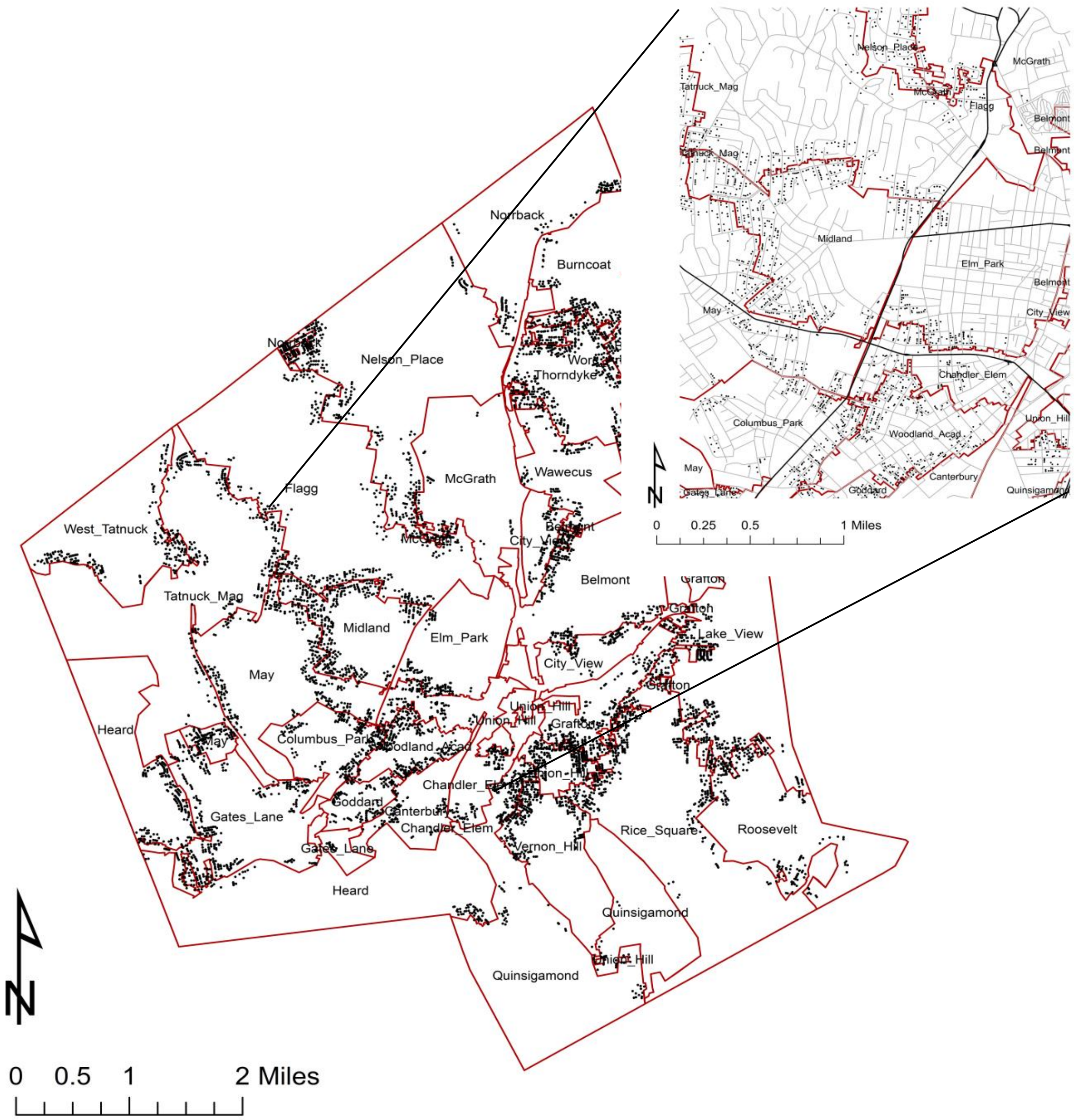
Rothstein, J. (2010). "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* **125**(1): 175-214.

Sanders, W. L. (2000). "Value-Added Assessment from Student Achievement Data: Opportunities and Hurdles CREATE NATIONAL EVALUATION INSTITUTE July 21, 2000." *Journal of Personnel Evaluation in Education* **14**(4): 329-339.

Tekwe, C. D., R. L. Carter, et al. (2004). "An empirical comparison of statistical models for value-added assessment of school performance." *Journal of Educational and Behavioral Statistics* **29**(1): 11-36.

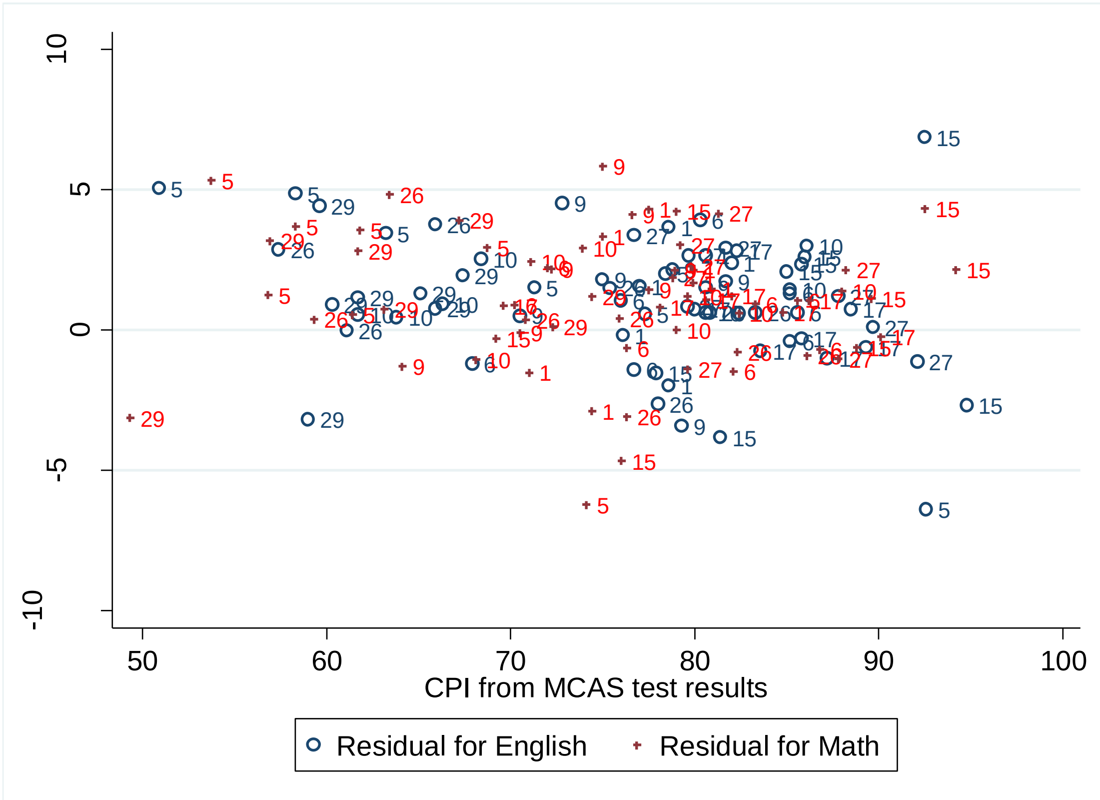
Tiebout, C. M. (1956). "A Pure Theory of Local Expenditures." *Journal of Political Economy* **64**(5): 416-424.

Figure 1: Elementary School Attendance Zones and Property Sales within 0.1 Miles of Boundaries



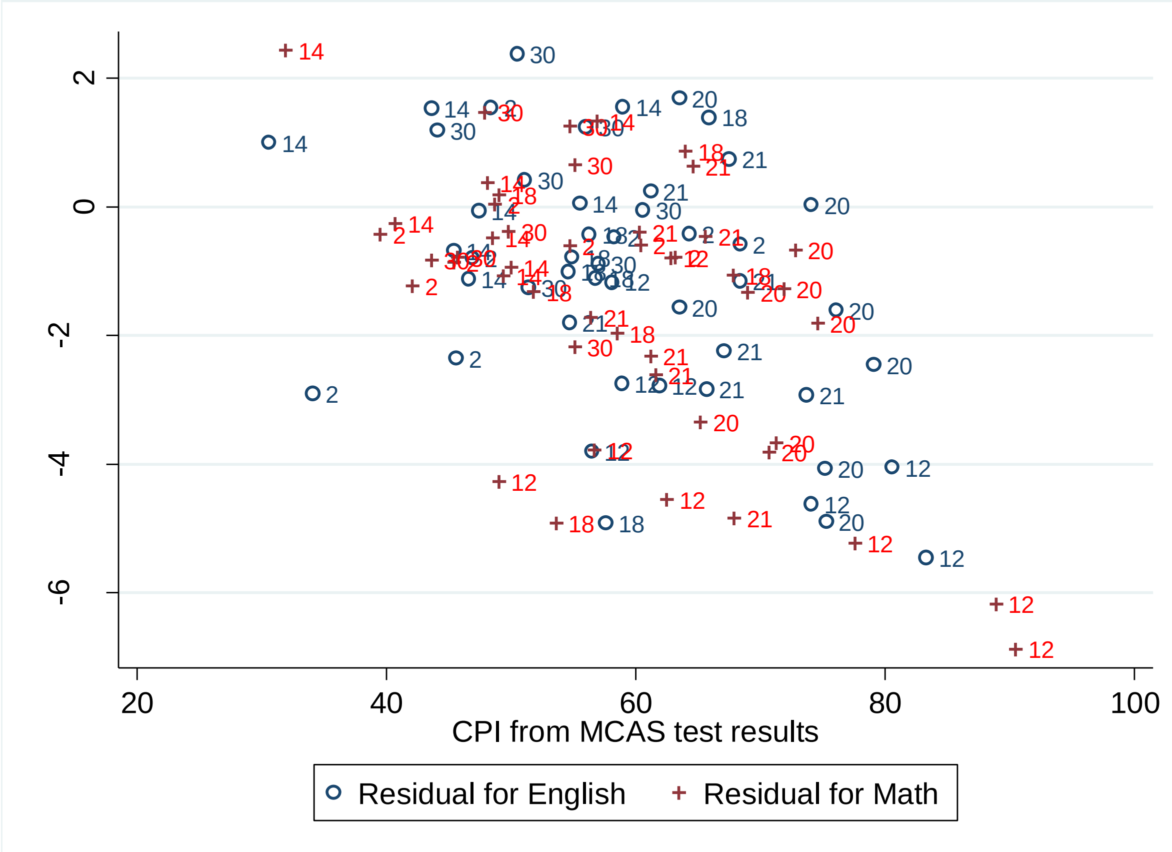
Notes: Attendance zones are presented in red. Each dot represents the sale of a one-, two- or three-family house within 0.1 miles of attendance zone boundaries during 2006-2011.
Source: Worcester Public Schools for attendance zone definitions and Warren Group for property sales data.

Figure 2: A comparison of MCAS test scores and Value-Added measures for high-performing schools



Notes: The residuals are from a regression of estimated value-added for each school-year pair on the CPI from test results. Only high-performing schools are included. For further explanation, please see the text.

Figure 3: A comparison of MCAS test scores and Value-Added measures for low-performing schools



Notes: The residuals are from a regression of estimated value-added for each school-year pair on the CPI (average performance index). Only low-performing schools are included. For further explanation, see the text.

Table 1: Summary Statistics of the Worcester Housing Sales Sample for 2006-2011

	Full Sample			0.10 mile Boundary Sample		
Variables	No of Obs.	Mean	St Dev.	No of Obs.	Mean	St Dev.
Log of Real House Price	26812	12.04	0.52	11868	12.02	0.54
Bathrooms	26812	1.96	0.8	11868	2.09	0.81
Bedrooms	26772	4.14	2.16	11852	4.6	2.39
Lot size (in square feet)	26811	2.93	6.47	11868	2.68	5.52
Age (in years)	26619	69.34	37.78	11797	73.6	39.16
Porch[c]	26812	0.016	0.13	11868	0.02	0.13
Outbuild[c]	26812	0.36	0.48	11868	0.33	0.47
Attic[c]	26812	0.14	0.35	11868	0.14	0.35
Air[c]	26812	0.06	0.24	11868	0.06	0.24
Distance to Boundary[d]	26812	0.133	703.08	11868	0.046	151.7

Notes: House prices are deflated to 2005 dollars using Consumer Price Index (CPI) data.
Sources: House sales prices and attributes are from The Warren Group. The CPI index is for the Boston metropolitan area.

Table 2: Summary Statistics of Student Test Score Data for 2006-2011

Variables	Mean	Standard Deviation
English Scaled Score	236.10	14.84
Math Scaled Score	232.00	17.47
Free/Reduced Lunch	0.63	
Limited English Proficiency	0.16	

Source: Adminstrative records of the Worcester Public Schools for MCAS test results for the period 2006-2011.

Table 3: The impact of schools on student achievement: School Value-added models

	FGLS		MLE		HLM	
	English	Math	English	Math	English	Math
Lagged Test score	0.64	0.69	0.82	0.86	0.77	0.82
	(166.47)	(183.06)	(282.28)	(309.73)	(234.95)	(270.41)
Limited English Proficiency	-0.14	-0.04	-0.03	0.01	-0.06	0.005
	(-13.24)	(-4.20)	(-4.86)	(2.64)	(-7.62)	(0.75)
Free/Reduced Eligible	-0.13	-0.12	-0.07	-0.06	-0.10	-0.08
	(-16.24)	(-15.01)	(-10.53)	(-10.25)	(-13.60)	(-12.29)
Grade 5	0.65	0.69	0.68	0.70	0.67	0.70
	(4.64)	(5.32)	(5.20)	(6.07)	(5.11)	(6.04)
Grade 6	0.65	0.68	0.68	0.70	0.67	0.70
	(4.63)	(5.25)	(5.17)	(5.99)	(5.09)	(5.97)
Grade 7	0.24	0.58	0.33	0.56	0.44	0.54
	(1.50)	(3.90)	(2.30)	(4.12)	(3.30)	(4.52)
Grade 8	0.23	0.57	0.35	0.56	0.45	0.53
	(1.48)	(3.83)	(2.38)	(4.11)	(3.37)	(4.44)
Grade 10	0.39	0.88	0.51	0.87	0.59	0.72
	(2.36)	(5.97)	(3.33)	(6.56)	(4.36)	(6.02)
Constant	-0.01	-0.49	-0.17	-0.49	-0.56	-0.62
	(-0.07)	(-3.29)	(-1.13)	(-3.61)	(-4.23)	(-5.25)
School Effects (SD)	0.13	0.16	0.10	0.13	0.15	0.16
Adjusted R-Squared	0.67	0.73				
N	38,370	34,680	38,370	34,680	38,370	34,680
N Schools	35	35	35	35	35	35

Notes: *t*-statistics in parentheses. FGLS and MLE value-added models estimate school effects or school value-added as fixed effects. FGLS specifies student unobserved heterogeneity as random effects (it is a random effects model). MLE structures individual errors as a first order Auto-regressive process (AR 1). For a discussion of the HLM (Hierarchical Linear Models), please see the text.

Table 4: Correlation matrix of different measures of school quality for Worcester Elementary Schools (2006-2011)

	Level Measure			Value-added Measures								
	Composite Performance Index (CPI)			Feasible Generalized Least Squares (FGLS)			Maximum Likelihood Estimates (MLE)			Hierarchical Linear Model(HLM)		
	English	Math	Average	English	Math	Average	English	Math	Average	English	Math	Average
Composite Performance Index (CPI)												
English	1											
Math	0.82	1										
Average	0.95	0.95	1									
FGLS												
English	0.37	0.38	0.39	1								
Math	0.37	0.41	0.41	0.69	1							
Average	0.40	0.43	0.43	0.89	0.94	1						
MLE												
English	0.24	0.25	0.26	0.92	0.61	0.80	1					
Math	0.25	0.30	0.29	0.62	0.94	0.87	0.62	1				
Average	0.27	0.31	0.30	0.83	0.89	0.94	0.86	0.93	1			
HLM												
English	0.30	0.31	0.32	0.97	0.64	0.84	0.97	0.60	0.84	1		
Math	0.30	0.34	0.33	0.65	0.96	0.90	0.62	0.97	0.92	0.62	1	
Average	0.34	0.37	0.37	0.86	0.91	0.97	0.85	0.91	0.98	0.86	0.93	1

Notes: The value-added measure for each school and year is derived from the regression results reported in Table 3. The measures have all been subject to empirical Bayes' adjustment

Table 5: Stability of School Performance over the Period 2006-2011

Performance Level	**Share of Schools attaining the Designated Performance Level**					
Number of years in Top Quartile (out of six years)	**Composite Performance Index (CPI) (Level Measure)**	**Feasible Generalized Least Squares (FGLS)**	**FGLS (Empirical Bayes)**	**Maximum Likelihood Estimates (MLE)**	**MLE (Empirical Bayes)**	**Hierarchical Linear Model (HLM)**
Never	50	48.57	51.43	36.67	40	43.33
1 Year	10	20	11.43	23.33	20	13.33
2 Year	10	5.71	14.29	13.33	13.33	16.67
3 Year	16.67	5.71	2.86	6.67	3.33	3.33
4 Years and over	13.33	20	20	20	23.33	23.33

Notes: The value-added measure for each school and year is derived from the regression results reported in Table 3. The empirical Bayes adjustment accounts for measurement error in school value-added estimates. The Hierarchical Linear Model (HLM) has an empirical Bayes adjustment built-in.

Table 6: Impact of the Current Year Value-Added (VA) Measures of Test Scores (Empirical Bayes' Maximum Likelihood Estimates) on Sales Prices

Independent Variable	VA All Students			VA estimated for Stayers only			VA estimated for Movers only		
	English (1)	Math (2)	Average (3)	English (4)	Math (5)	Average (6)	English (7)	Math (8)	Average (9)
Scaled test score	0.0120	0.00121	0.00614	0.0148	0.00201	0.0120	0.0115	-0.000447	0.00462
	(1.79)	(0.25)	(0.89)	(1.61)	(0.19)	(0.92)	(1.54)	(-0.08)	(0.60)
Baths	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	(9.53)	(9.51)	(9.57)	(9.52)	(9.53)	(9.58)	(9.46)	(9.40)	(9.45)
Rooms	0.0399	0.0399	0.0399	0.0400	0.0399	0.0399	0.0400	0.0399	0.0399
	(6.87)	(6.76)	(6.80)	(6.87)	(6.76)	(6.82)	(6.84)	(6.77)	(6.80)
Lotsize (1000 square feet)	0.00608	0.00604	0.00606	0.00608	0.00603	0.00606	0.00605	0.00603	0.00604
	(4.46)	(4.45)	(4.44)	(4.49)	(4.45)	(4.46)	(4.43)	(4.45)	(4.44)
Age	-0.00225	-0.00226	-0.00225	-0.00225	-0.00226	-0.00225	-0.00226	-0.00227	-0.00226
	(-8.12)	(-8.15)	(-8.09)	(-8.19)	(-8.13)	(-8.09)	(-8.18)	(-8.21)	(-8.20)
	0.110	0.109	0.109	0.110	0.109	0.110	0.108	0.108	0.109
Air Conditioning	(4.02)	(3.98)	(4.00)	(4.00)	(3.97)	(3.98)	(4.01)	(3.97)	(4.00)
	11.54	11.50	11.51	11.54	11.50	11.52	11.50	11.50	11.50
Constant	(259.52)	(257.81)	(257.24)	(221.82)	(260.83)	(239.42)	(239.88)	(261.56)	(257.80)

Observations	4500	4500	4500	4500	4500	4500	4500	4500	4500
Adjusted R-squared	0.37	0.36	0.37	0.37	0.36	0.37	0.37	0.36	0.37

Source: Results of ordinary least squares regression analysis of property sales data from the period 2006-2013.

Notes: t-statistics are in parentheses. The dependent variable is the log of the sales price. The value-added measure for each school and year is derived from the regression results reported in Table 3. Additional controls are included for the year of sale and boundary fixed effects. The results are for buffers of about 0.1 mile on either side of the attendance zone boundary. Stayers are defined as students who attended the same elementary school and the same middle school during their residence in Worcester. Movers are students who either changed elementary or middle schools or who entered the system after the third grade. Non-empirical Bayes' estimates are the estimates of school effects that are not adjusted for measurement error. All specifications have boundary fixed effects.

Table 7: Impact of the Current Year Value-Added Measures of Test Scores (HLM Estimates) on Sales Prices

Independent Variable	VA All Students			VA estimated for Stayers only			VA estimated for Movers only		
	English (1)	Math (2)	Average (3)	English (4)	Math (5)	Average (6)	English (7)	Math (8)	Average (9)
Scaled test score	0.00794	0.000951	0.00441	0.00671	0.000927	0.00405	0.00688	-0.000292	0.00338
	(1.65)	(0.24)	(0.83)	(1.46)	(0.22)	(0.73)	(1.58)	(-0.07)	(0.67)
Baths	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	(9.43)	(9.38)	(9.44)	(9.42)	(9.38)	(9.44)	(9.35)	(9.29)	(9.35)
Rooms	0.0400	0.0399	0.0399	0.0400	0.0399	0.0399	0.0400	0.0400	0.0399
	(6.86)	(6.76)	(6.80)	(6.87)	(6.76)	(6.80)	(6.83)	(6.77)	(6.80)
Lotsize (1000 square feet)	0.00608	0.00605	0.00607	0.00608	0.00605	0.00607	0.00605	0.00604	0.00606
	(4.45)	(4.44)	(4.44)	(4.47)	(4.45)	(4.45)	(4.43)	(4.44)	(4.44)
Age	-0.00225	-0.00225	-0.00225	-0.00225	-0.00225	-0.00225	-0.00226	-0.00226	-0.00226
	(-8.10)	(-8.10)	(-8.06)	(-8.12)	(-8.07)	(-8.02)	(-8.18)	(-8.20)	(-8.19)
Air Conditioning	0.110	0.109	0.110	0.110	0.109	0.110	0.109	0.109	0.109
	(4.02)	(4.00)	(4.01)	(4.00)	(3.99)	(4.00)	(4.02)	(3.99)	(4.01)

Constant	11.44	11.47	11.45	11.46	11.47	11.46	11.44	11.47	11.46
	(247.28)	(290.64)	(257.13)	(288.47)	(287.29)	(274.74)	(215.19)	(298.65)	(257.89)
Observations	4500	4500	4500	4500	4500	4500	4500	4500	4500
Adjusted R-squared	0.37	0.36	0.36	0.37	0.36	0.36	0.36	0.36	0.36

Source: Results of ordinary least squares regression analysis of property sales data from the period 2006-2013.

Notes: *t*-statistics are in parentheses. The value-added measure for each school and year is derived from the regression results reported in Table 3. The dependent variable is the log of the sales price. Additional controls are included for the year of sale. Stayers are defined as pupils who attended the relevant elementary school for four years or more. Movers are pupils who attended the school for only one year. Non-empirical Bayes' estimates are the estimates of school effects that are not adjusted for measurement error. All specifications have boundary fixed effects, and the sample is restricted to houses within 0.10 miles of school attendance zone boundaries.

Table 8: Treatment Effect of High Value-added School on House Prices and Falsification Test

Independent Variable	Actual School Catchment Boundary	Artificial School Catchment Boundary
High Side Dummy	0.0236	-0.0147
	(1.39)	(-1.05)
Baths	0.121	0.143
	(9.72)	(8.39)
Rooms	0.0405	0.0219
	(7.37)	(3.33)
Lotsize (1000 square feet)	0.00614	0.00346
	(4.46)	(3.04)
Age	-0.00232	-0.00199
	(-8.42)	(-6.39)
Air Conditioning	0.108	0.164
	(3.96)	(6.01)
Constant	11.49	11.48
	(238.76)	(209.43)
Observations	4500	4342
Adjusted R-squared	0.37	0.35

Source: Results of ordinary least squares regression analysis of property sales data from the period 2006-2013.

Notes: t-statistics are in parentheses. The key explanatory variable is a dummy variable (high side dummy) which indicates whether a house is on the side of the catchment boundary that has the school with higher value-added measure (for any given year) than the same for the school across the boundary. The dependent variable is the log of the sales price. Additional controls are included for the year of sale. All specifications have boundary fixed effects, and the sample is restricted to houses within 0.10 miles of the artificial school attendance zone boundaries, or within 0.10 miles of the actual boundaries.

Table 9: Impact of the Student Growth Percentile (SGP) on Sales Prices

Independent Variable	(1)	(2)	(3)	(4)	(5)
Student Growth Percentile (SGP) - English	0.00158	0.00138	0.000639	0.000615	0.000615
	(2.18)	(1.92)	(0.98)	(0.88)	(0.88)
Estimated Value-Added - English		0.00849	0.00814	0.0114	0.0114
		(1.15)	(1.28)	(1.52)	(1.52)
Baths	0.115	0.112	0.103	0.101	0.101
	(6.44)	(6.62)	(6.20)	(6.18)	(6.18)
Rooms	0.0366	0.0375	0.0415	0.0406	0.0406
	(5.72)	(5.75)	(6.60)	(6.54)	(6.54)
parcel_size2	0.00463	0.00563	0.00591	0.00555	0.00555
	(2.81)	(4.05)	(3.99)	(3.64)	(3.64)
Age	-0.00239	-0.00255	-0.00237	-0.00241	-0.00241
	(-7.35)	(-7.52)	(-7.10)	(-6.99)	(-6.99)
Air	0.122	0.129	0.131	0.127	0.127
	(3.37)	(3.48)	(3.68)	(3.62)	(3.62)
Constant	11.94	11.87	10.68	14.37	10.75
	(69.04)	(124.92)	(87.78)	(84.75)	(63.84)
PSA Dummies	No	No	No	Yes	Yes
Census Block Group Dummies	No	No	Yes	No	Yes
Observations	3052	2883	2883	2883	2883
Adjusted R-squared	0.31	0.31	0.34	0.34	0.34

Source: Results of ordinary least squares regression analysis of property sales data from the period 2006-2013.

Notes: t-statistics are in parentheses. The value-added measure for each school and year is derived from the regression results reported in Table 3. The dependent variable is the log of the sales price. Additional controls are included for the year of sale. All specifications have boundary fixed effects, and the sample is restricted to houses within 0.10 miles of school attendance zone boundaries.