The Capitalization of School Quality:

Evidence from San Diego County

Christopher Harris

Spring 2001

CHAPTER I

INTRODUCTION

Over 40 years ago, Charles Tiebout (1956) suggested that households shop around the different communities in a metropolitan area for the community that offers the bundle of public goods that best suits them. The public good of most interest to families with children is perhaps the public school system provided by a community. There are typically many school districts located within a metropolitan area, providing families with a wide variety of choice. Each school district differs in the quality of education it provides, and a family reveals its preference for school quality through its choice of housing location.

Real estate agents respond to parental concerns about public school quality by providing house buyers with information about the schools and school districts located in different communities. Similarly, metropolitan newspapers often publish the test scores of students enrolled in local public schools as a service to those in the metropolitan area. If families respond to this information as Tiebout proposed they might, then the price of a house should not only reflect its physical characteristics, but also the quality of public education provided by the local district. In other words, homes should have a premium attached to their price that reflects the quality of the local public schools.

How much are parents willing to pay for a house located in a good school district? Wallace Oates (1969) offered the first empirical test of the Tiebout hypothesis. Using data from 53 northeastern New Jersey municipalities, Oates found that a $100 increase in expenditure per pupil led to a $1,200 increase in housing values, all else equal. Since Oates' work, several other studies have examined the relationship between housing values and public school quality.[1] Most of these studies use class size, expenditure per pupil, or standardized test scores as a measure of school quality. These studies generally found that improvements in school quality led to higher housing prices, all else equal.

In this paper, I provide two important contributions to the literature on school quality capitalization. First, I examine the relationship between student performance on standardized tests and housing values in San Diego County, a major metropolitan area that is yet to be studied in this manner. The measure of school quality used is the combined average of district-level fourth and tenth grade math and reading scores. Second, I build on the existing literature on school quality capitalization by examining whether school district housing price premiums vary systematically with the structural and neighborhood characteristics associated with a home. Specifically, theory suggests that in heterogeneous communities, the impact school quality has on housing values will depend on the structural and neighborhood characteristics of a home. For example, as noted by Sonstelie and Portney (1978), households with children place a higher value on homes located in good school districts relative to households without children. Those same households with children most likely also place a higher value on larger homes relative to households without children. Thus, if public school quality within a district increases, the value of large homes should increase more than proportionately to the value of small homes.

Previous studies examining school quality capitalization have ignored the importance that these interaction effects have on consumer location choice. I contribute to the school quality capitalization literature by providing a more accurate explanation of how consumers

[1] See Bergstrom et al (1982), Black (1999), Brunner et al (2000), Crone (1998), Figlio (2000), Jud and Watts (1981), Reinhard (1981), and Sonstelie and Portney (1980).

in the housing market choose their housing locations, and test whether housing price premiums within a school district vary systematically with the structural and neighborhood characteristics of a home.

To test the hypothesis that school quality is capitalized into housing values, I collected data from the 1998 Stanford 9 (SAT 9) exam and home sales in San Diego County occurring between June 1998 and May 1999. I find that increases in the average combined math and reading score did significantly increase housing values; a one-point increase in average score increases housing values 0.2%, all else equal. I also find that the interactions between school quality, house size, and the size of the senior citizen population significantly affect housing prices.

CHAPTER II

LITERATURE REVIEW

Tiebout (1956) developed the underlining theory used in my paper. He proposed that homebuyers shop for public goods just as they shop for a home in which to enjoy them. Where Musgrave (1939, 1955) and Samuelson (1954, 1955) made the assumption that the federal government handles expenditures on public goods, Tiebout argued that local governments provide police, courts, and education. "Consumer-voters", as he called homebuyers, choose homes in a community whose local government offers the package of public goods that best satisfies their wants. Consumers' demand for public goods determines where they live. Possibly the public good of greatest importance to families with children is public education. Large metropolitan areas typically have several school districts within their boundaries, which allow consumers to shop for public education.

In support of the Tiebout Hypothesis, Crone (1998) points out that prospective homebuyers are provided with information to aid their decision-making process. Crone mentions that newspapers occasionally publish information related to school quality, such as expenditure per student, student/teacher ratios, and standardized test scores. For example, the July 19, 2000 issue of *The San Diego Union-Tribune* included a special section that provided SAT 9 test scores from every school in San Diego. Similarly, the October 5, 2000 issue of *The San Diego Union-Tribune* published the rankings of San Diego County schools based on SAT 9 scores. Crone also states that real estate agents provide information about local public schools on their house listings. This is a common practice in many parts of San Diego County. As Crone showed, homebuyers have the information necessary to shop for public education. If buyers shop for public education while shopping for a house, then house prices will reflect this. The price of a house will include the value of the structural characteristics and neighborhood characteristics, as well as a premium for the quality of public education.

How much are parents willing to pay for a home in a high-quality school district? If all houses had the same structural and neighborhood characteristics, and were located in the same public school district, then housing prices would be the same. If houses only varied by school district attendance, the effect that school quality has on housing prices would be easy to determine; the difference in housing prices would be due only to differences in school quality. This difference is the school premium. Unfortunately, houses very seldom have the same structural and neighborhood characteristics, so prices vary for reasons other than differences in school premiums. As Crone (1998) points out, the structural and neighborhood characteristics of a house must be separated from the school premium in order to find the effect that school quality has on housing prices.

Papers that empirically examine how housing prices are affected by school quality, separate from the effects of structural and neighborhood characteristics, are applications of hedonic price theory.[2] Hedonic price theory states that the value of a heterogeneous good equals the combined value of the good's characteristics. The value of a house, for example, is equal to the value of the house's structural characteristics, neighborhood characteristics, and local public school quality. Multivariate regression analysis separates the effects that individual characteristics have on the total value of a heterogeneous good. Therefore, the influence each individual characteristic has on the total value can be determined. Accounting for the influence of an individual characteristic separate from the other characteristics allows for the "all else equal" statement to be made.

Wallace Oates (1969) provided one of the most influential studies examining the effect that public school quality has on housing values. His study examined the relationship between housing values and inter-jurisdictional differences in public school quality in northern New Jersey. Oates used expenditure per pupil as a proxy for public school quality. In order to determine the value of a house, he used the two-stage least squares (TSLS) estimation procedure on a model that accounted for the local property tax rate, expenditure per pupil, and various structural and neighborhood characteristics associated with a house. TSLS was employed to eliminate the simultaneity bias stemming from the relationship between median housing values and expenditure per pupil. Simultaneity bias occurred in his

[2] See Bergstrom and Goodman (1973), Rosen (1974), Bergstrom, Rubinfeld, and Shapiro (1982), and Cassell and Mendelsohn (1985).

sample because housing values dictate the property tax rate, and the resulting property tax revenues determined the level of expenditure per pupil in that local school district. The expenditure per pupil in that school district then affects the housing values in that neighborhood. In short, housing values determine expenditure per pupil, which then determines housing values. TSLS alleviated this problem. Oates concluded that increasing expenditure per pupil from $350 to $450 caused housing values to increase $1,200, all else equal.

Oates' work has inspired many others to investigate the relationship between housing values and public school quality, each using samples from different metropolitan areas and different measures of school quality. For example, Reinhard (1981) claimed further improvement could be made to Oates' model. Reinhard developed a dependent variable for his model to include the current discount rate. Reinhard applied his model, and Oates' original model, to a sample of data from San Mateo County, California. Reinhard's model estimated that a one-month improvement in reading level led to a $1,468 increase in house price, and a one dollar increase in expenditure per student led to a $30 increase in house price, all else equal.

Sonstelie and Portney (1980) used gross rent, rather than a house's sale price, as the dependent variable in their hedonic model. They argued that house prices in some areas will not capitalize school quality, but gross rents will. Another claim is that the difference in property tax payments is capitalized, not the difference in property tax rates. The difference is that property tax rates do not affect a consumer's choice, but the difference in the property tax bills of two or more houses (and thus the difference in expenditure per pupil) will affect the consumer. The consumer will choose the house with a smaller property tax bill, or the house that receives a larger bundle of public goods (higher school quality) in return for the higher property tax bill. For their sample of houses in San Mateo County, California, Sonstelie and Portney found that the annual gross rent increases $52 for the median house when there is a marginal improvement in reading ability. A $1 increase in expenditure per pupil increases gross rent by $0.90 for the median house.

Jud and Watts (1981) looked at the effect that both school quality and racial composition had on housing prices. Their data was from Charlotte, North Carolina, and looked at housing prices shortly after court-ordered school desegregation took place. Using a semi-log model, they initially found that racial composition (the proportion of black students in a school) hurt housing prices. After including reading achievement test scores in their model, they saw that racial composition became statistically insignificant. Their conclusion was “that an increase of one grade level in the achievement level of neighborhood schools is associated with an increase of 5.2% to 6.2% in the value of the average house” (p. 467).

Black (1999) studied neighborhoods in suburban Boston. In addition to state standardized test scores (sum of the average reading and math scores), she used expenditure per student, student/teacher ratio, and the existence of reduced-cost or free preschool programs as measures of school quality. She points out that a major problem affecting school-focused housing price studies is that conventional methods do not control for all the neighborhood characteristics that differ between school districts. These omitted variables lead to biased results. To correct this, Black replaces the variables measuring neighborhood and school district characteristics with a series of dummy variables, which designate houses that share school district boundaries. The dummy variables control for the unobserved house and neighborhood characteristics that houses on opposite sides of school district boundaries share, thus preventing omitted variables. Accounting for these similarities, she found that buyers were willing to pay 2.1% more for a house near schools that scored 5% higher than the mean score on state tests. This was equivalent to a $3,948 difference in price for the average house. She also estimated that a one point increase in average reading and math scores could lead to a 1.5% increase in average housing prices in the state of Massachusetts.

Figlio and Lucas (2001) are working on a study closely related to this paper. Their study looks at the school grading system that the Florida Department of Education instituted in 1999. A school is graded by student performance on various Florida state standardized tests, the percentage of its students that took the exam, the school's absentee rate, and its suspension rate. Based on these criteria, each school received a grade "A" through "F".

Figlio and Lucas (2001) focus on Gainesville, FL. There is only one school district in that city, but the district is divided into several neighborhoods. Controlling for student performance on state tests, they found that buyers were willing to pay $9,179 more for houses near schools graded "A" than for houses near schools graded "B". They found that houses in the relatively more-expensive neighborhoods increased in value after the local schools received an "A", while houses in similar neighborhoods decreased in value after the local schools received a "B". Similar trends were discovered for houses in relatively less-expensive neighborhoods.

CHAPTER III

THEORETICAL FRAMEWORK

Consumers choose housing locations that offer the combination of structural characteristics, neighborhood characteristics, and public goods that most closely match their preferences. These locations implicitly reveal each household's desired level of quality. A household receives utility from its consumption of goods, including the quality of housing it purchases. Housing itself is a function of three sets of variables:

$$H = f(S, N, Q)\,, \tag{1}$$

where S denotes the house's structural characteristics, such as square footage, number of bathrooms, the presence of a pool, etc.; N denotes the house's neighborhood characteristics, such as the distance to the beach, the distance to work, and crime rates; and Q denotes the quality of local public schools.

There are three underlying assumptions about the housing market. First, urban areas can be treated as a single housing market. Individuals shopping in the market must be perfectly informed about all of their alternatives, and should be free to choose a housing location anywhere in the urban area. If individuals can shop freely throughout the entire urban area, they can choose a house that contains the bundle of characteristics that suits them best. Second, enough variation exists in the housing market to allow households to adjust the quantity of any one characteristic, and still be able to find an alternate location similar in every respect except the quantity offered of that one characteristic. Third, the housing market is in a competitive equilibrium.

At a competitive equilibrium, households are equally well off paying less for a house with fewer desirable characteristics as they are paying more for a house with more desirable characteristics. Knowing that households receive utility from their consumption of housing, households can maximize their utility by choosing the house with the most desired bundle of characteristics, given the price. Utility maximization is each household's ultimate goal.

Prior studies have examined the school quality capitalization hypothesis using the following specification:

$$P_i = \beta_0 + \beta_1 S_i + \beta_2 N_i + \beta_3 Q_i + \varepsilon_i \,, \qquad (2)$$

where P_i denotes either the sale price of a home or the natural log of the sale price of a home, and ε_i is a random disturbance term. In equation (2), the implicit price of an additional unit of school quality is simply β_3. Thus, equation (2) assumes that the impact school quality has on housing values is independent of the structural and neighborhood characteristics of a home.

However, as noted by Sonstelie and Portney (1978), in heterogeneous communities, an increase in public school quality may have different effects on the value of large and small homes. In particular, if public school quality within a community increases, the value of large homes should increase more than proportionately to the value of small homes (Sonstelie and Portney, 1978 p. 272). More generally, when the housing stock in a community is not homogeneous, the impact a change in public school quality has on the value of a home will depend on the structural and neighborhood characteristics of the home.

To illustrate that point, consider an area comprised of two communities. The communities are identical except for the quality of their schools. Community 1 has high-quality schools and Community 2 has low-quality schools. Furthermore, both Community 1 and Community 2 contain a heterogeneous housing stock. In particular, Community 1 and Community 2 both contain large and small homes. The communities are illustrated in Figure 1.

As Figure 1 shows, four housing options exist: low-quality/small homes, high-quality/small homes, high-quality/large homes, and low-quality/large homes. The demand for these various housing bundles will depend on the characteristics of households in the area. For simplicity, assume there are just two types of households: households with children and households without children. Households with children are assumed to demand only large homes, whereas households without children demand only small homes. Thus, in terms of Figure 1, households with children locate in quadrants III and IV. In contrast, households without children locate in quadrants I and II.

Because households without children do not demand school quality, they are indifferent between homes located in the high-quality community and homes located in the low-quality community. As a result, they would be unwilling to pay more for a small home in the high-quality community. Thus, small homes should cost the same, regardless of the type of community (high-quality schools or low-quality schools) in which they are located. This suggests that homes located in quadrants I and II should have the same price.

Households with children demand large homes; they also demand housing in high-quality districts. Therefore, there should be premium attached to large homes located in the high-quality community. In particular, large homes in Community 1 should command a higher price than large homes in Community 2.

Although simplistic, this example illustrates an important point: in heterogeneous communities, the effect that inter-jurisdictional differences in public school quality has on the value of a home depends on the structural and neighborhood characteristics of the home. In the simple example outlined above, inter-jurisdictional differences in public school quality will have no effect on the price of small homes. In contrast, inter-jurisdictional differences in public school quality will be capitalized into the value of large homes.

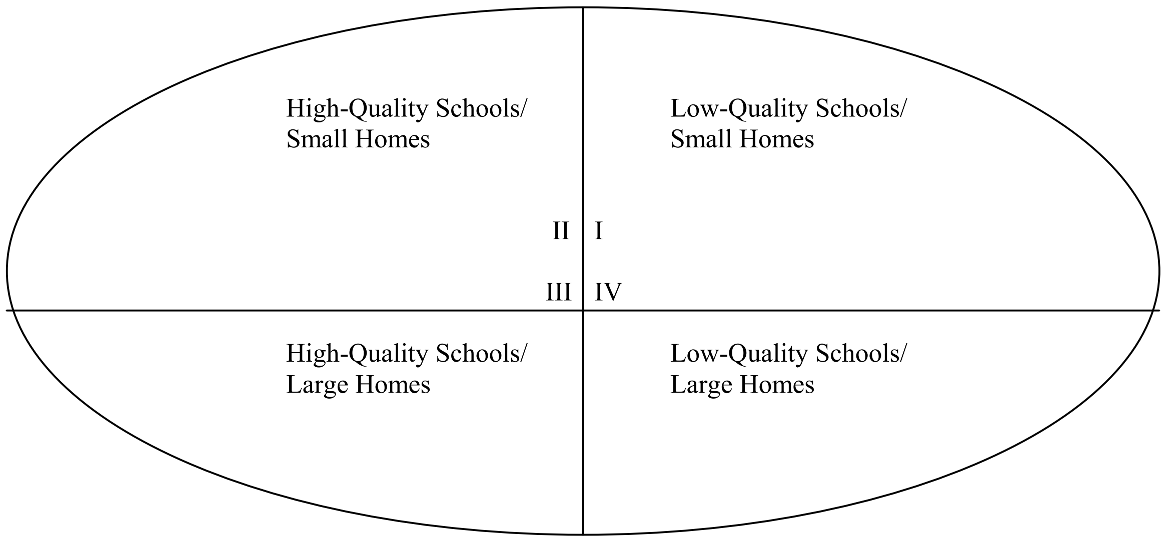

Figure 1

A Diagram of Housing Options

The theory outlined above suggests that the impact public school quality has on the value of a home depends on the structural characteristics of the home. This suggests that the housing price function given in equation (2) is specified incorrectly. In particular, theory suggests that the correct specification is

$$P_i = \beta_0 + \beta_1 S_i + \beta_2 N_i + \beta_3 Q_i + \beta_4 (Q_i * S_i), \qquad (3)$$

where $S_i * Q_i$ is the interaction between house size and school quality. Previous studies ignored the inclusion of such interactions, suggesting that those models were incorrectly specified. In terms of equation 3, the implicit price of an additional unit of school quality is now

$$\frac{\partial P_i}{\partial Q_i} = \beta_3 + \beta_4 S_i . \qquad (4)$$

As can be seen from equation 4, the effect school quality has on housing values now depends on the structural characteristics of a home. To illustrate that point, consider once again the two communities illustrated Figure 1, where homes are either small or large. Let S_i equal zero if a home is small and one if a home is large. According to the theory outlined above, the price of small homes should be independent of school quality, implying that:

$$\frac{\partial P_i}{\partial Q_i} = \beta_3 + \beta_4 * 0 = 0 . \tag{5}$$

In contrast, large homes in good school districts command a premium, implying that:

$$\frac{\partial P_i}{\partial Q_i} = \beta_3 + \beta_4 * 1 > 0 . \tag{6}$$

Equations (5) and (6) also illustrate another important point: the theory outlined above suggests that β_3 should be zero. In other words, in heterogeneous communities, school quality has no direct impact on housing values – there is only an indirect effect through the interaction of school quality with the structural characteristics of a home. Thus, consider a regression model where house price is regressed on the structural characteristics of a home, neighborhood characteristics, school quality (as measured by test scores) and an interaction between house size and school quality. The theory outlined above suggests that the coefficient on school quality should be zero while the coefficient on house size interacted with school quality should be positive.

Of course the theory outlined above is based on a number of strong assumptions. For example, homes are neither simply small nor large. Homes of a variety of different sizes exist, and no clear line differentiating between "small" and "large" exists. Furthermore, the demand for large and small homes is not determined solely by the presence or absence of children in a household (as was assumed above). In reality, some households without children demand large homes and some households with children demand small homes. These facts suggest that in reality, the impact school quality has on housing values will not be as clear cut as theory suggests. Nevertheless, the theory outlined above makes clear an important point: in the presence of heterogeneous communities, to properly estimate the impact school quality has on housing values, one must allow school quality to affect housing values indirectly through interactions with the characteristics of a home and neighborhood.

CHAPTER IV

DATA AND EMPIRICAL SPECIFICATION

The data used in this study come from several sources. The housing data comes from the Experian Company. Each observation represents the sale of an owner-occupied, single-family house sold in San Diego County between June 1998 and May 1999. There are 74,486 observations.[3] Variables describing the structural characteristics of a home include: (1) house size, measured by the square footage of living space, (2) presence of a pool, and (3) presence of a view.

I utilize 5 measures of neighborhood quality/community quality. Four of those measures were constructed using data from the *1990 Census of Population and Housing*. These neighborhood variables are measured at the census tract level and include: (1) the

[3] Some restrictions were made to the used data. First, houses that sold for less than $20,000 were not included in the sample, for they may not represent true market transactions. Second, houses that sold for less that $35 per square foot or more than $1,000 per square foot were not included in the sample.

percent of population white, (2) the percent of population below poverty, (3) percent of population age 65 or older, and (4) travel time to work. In addition to these measures, I utilize one other measure of neighborhood quality. This variable identifies homes located along the coast. In particular, the dummy variable *Coast* has a value of one if the home sale is located in a zip code that lies along the San Diego County coastline. Cities with coastal locations are Oceanside, Carlsbad, Encinitas, Solana Beach, Del Mar, Coronado, and Imperial Beach. Communities of the city of San Diego that have coastal locations are La Jolla, Pacific Beach, Mission Beach, Ocean Beach, and Point Loma. A total of 18 zip codes carried values of *Coast* equal to one. To measure school quality, I utilize district-level data on student performance on the 1998 Stanford Achievement Test, Ninth Edition (SAT 9). The SAT 9 was a standardized exam in math and reading administered to all fourth and tenth graders. The data was obtained from the California Department of Education.

San Diego County contains a diverse number of school districts. Within the county there are 26 elementary school districts, 6 high school districts, and 9 unified school districts. Because the boundaries of elementary and high school districts overlap (each high school district usually contains several elementary school districts), elementary school districts were combined with their appropriate high school districts, creating a total of 35 districts within the sample**.** To create an index of public school quality, I used the average 4^{th} and 10^{th} grade math scores and the average 4^{th} and 10^{th} grade reading scores. In particular, I summed the average math and reading scores for 4^{th} graders and the average math and reading scores for

10th graders, and then divided that sum by two. The result is a weighted index of district-level student performance.[4]

Several previous studies used standardized test scores as the measure of school quality, while others used expenditure per pupil. Thus, one might ask why I only utilize test scores as a measure of school quality. The answer is related to California's system of public school finance. Over that last 25 years, California has transferred the responsibility of financing its public schools from local school districts to the state, equalizing spending per pupil across districts in the process. As noted by Downes (1992) and Sonstelie et al (2000), by the early 1990's, very little variation remained in spending per pupil across districts. As a result, spending per pupil does not capture inter-jurisdictional differences in public school quality in California. However, Downes (1992) shows that despite the equalization of school funding, student achievement on standardized tests still varies widely across school districts in California. This suggests that standardized test scores are a better measure of school quality in cases involving California.

Table 1 contains the names and definitions all variables used in the study. Table 2 contains the means and standard deviations of those variables. The mean house sale price during the period June 1998 to May 1999 was $239,097. The average house was 1,784 square feet. Twenty-three percent of the houses lie in coastal communities. Neighborhoods with the highest proportion of senior citizens were in the Poway Unified and San Marcos Unified school districts; there are neighborhoods in both districts with more than 82% age 65 or older.

---

[4] School quality could also be measured using two separate variables based on the 4th and 10th grade test scores. However, due to the strong correlation between 4th and 10th grade scores (over .9) I decided to use a weighted average of those scores.

Rancho Santa Fe Elementary has the highest average combined math and reading score at 1393. The two largest school districts, San Diego Unified and Poway Unified, scored 1315 and 1371. San Ysidro Elementary has the lowest combined score of 1274. The average score for the entire sample was 1330.

The sale price of a house is estimated as a function of a house's structural characteristics, neighborhood characteristics, and public school quality. The estimation equation for house *i*, in census tract *j*, attending schools in district *k* is

$$\begin{aligned} \ln(SalePrice)_{ijk} = {} & \beta_0 + \beta_1 LivArea_{ijk} + \beta_2 Pool_{ijk} + \beta_3 View_{ijk} + \\ & \beta_4 Coast_j + \beta_5 White_j + \beta_6 Poverty_j + \beta_7 TimeToWork_j + \\ & \beta_8 Age65_j + \beta_9 Score_k + \gamma City + \mu_k + \varepsilon_{ijk} \end{aligned} \quad (6)$$

where *LivArea*, *Pool*, and *View* are house-specific variables, *Coast*, *White*, *Age65*, *Poverty*, and *TimeToWork* are neighborhood-level variables, *Score* is a district-level variable, *City* is a vector of city-specific fixed effects designed to capture all city-level characteristics, ε_{ijk} is the random disturbance term and μ_k is a district-specific random effect.

Table 1

List of Variables

Variable	Definition (units)
Structural Characteristics	
Livarea	Size of house (square feet)
Pool	Presence of pool (dummy variable)
View	Presence of view (dummy variable)
Neighborhood Characteristics	
Coast	Coastal location (dummy variable)
White	Percent of population that is white (%)
Poverty	Percent living below poverty (%)
Time to work	Travel time to work (minutes)
Age65	Percent of population age 65 or older (%)
School Characteristics	
Score	Combined average of fourth and tenth grade math and reading score on Stanford 9 exam
Interaction Terms	
Score*Livarea	Interaction between Score and Livarea
Score*age65	Interaction between Score and %age65
Dependent Variable	
ln(Sale Price)	Natural log of sale price

Equation (6) has several noteworthy features. First, an inherent problem with the estimation of the impact of local amenities, such as public school quality, on housing values is omitted variable bias. If the level of services provided by a community is correlated with some unobservable community characteristics, estimates of the impact of local public service provision on housing values will be biased. The presence of city-specific fixed effects in equation (6) mitigates the problem of omitted variable bias since it captures permanent differences between cities with city fixed effects. It therefore allows one to control for unobservable city-specific factors that may be correlated with test scores. Specifically, the city-specific fixed effects remove the influence that city characteristics may have on school quality. Because school district boundaries extend across city boundaries, different levels of school quality will exist within a single city. Thus, *Score* measures the effect that a change in school quality within a specific city has on housing prices within that specific city. Only three of the 24 cities within San Diego County contain a single school district: Imperial Beach, Ramona, and Solana Beach.

Second, the literature is not consistent in regards to the functional form used in hedonic price models. Some studies employ a linear model, while others choose a semi-log model. Sonstelie and Portney (1978) explain that linear models are only necessary when the allotment of a home's characteristics (structural, neighborhood, and school quality) are separable, and can be recombined at a later time. As this is not possible, they argue against the use of linear models (p. 265).

Table 2

Summary Statistics

Variable	Mean	Standard Deviation
Structural Characteristics		
Livarea (ft^2)	1,784	850
Pool	0.15	0.35
View	0.30	0.46
Neighborhood Characteristics		
Coast	0.23	0.42
White (%)	73.5	18.1
Poverty (%)	5.2	5.6
Time to work (minutes)	12.4	2.1
Age65 (%)	11.9	8.2
School Characteristics		
Score	1330	24.6
Dependent Variable		
SalePrice ($)	239,097	190,905

74,486 observations is sample.

Theory suggests that consumers prefer more of each household amenity, however, their willingness to pay for additional quantities increases at a decreasing rate. Given this situation, the semi-log functional form would be a logical choice. To justify my choice of functional form, I used the Box-Cox transform to check for the appropriate dependent variable. The Box-Cox transform is

$$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda}, \qquad (7)$$

where the functional form is determined from
$$y^{(\lambda)} = \begin{cases} y-1 & if \quad \lambda = 1 \\ \ln(y) & if \quad \lambda = 0 \\ 1-\dfrac{1}{y} & if \quad \lambda = -1 \end{cases}$$

If λ = 0, then the functional form should be semi-log. If λ = 1, then the model should have a linear functional form. If λ = −1, then a reciprocal functional form should be used. The Box-Cox transform of *Sale Price* versus the explanatory variables from equation (6) calculated λ = 0.15. This value of λ suggests that the semi-log functional form is best suited for my sample, reiterating what theory suggests.

Third, a major problem of concern stems from the presence of different houses located in the same school district. There are unobservable characteristics that will affect housing values common to all houses in a school district. These unobserved district-specific effects are represented by the district-specific error term, μ_k in equation (6). The presence of μ_k in equation (6) will induce correlation among houses in the same school district. Moulton (1986) shows that this type of group-wise heteroscedasticity leads to standard errors that are biased toward zero, inflating t-scores. Following Moulton, I utilize a random effects estimator to control for group-wise heteroscedasticity.

The random effects estimator is a type of Generalized Least Squares (GLS) transformation. For explanatory purposes, consider a basic model

$$\begin{aligned} y_{ijk} &= \beta_0 + \beta_1 X_{ijk} + \omega_{ijk} \\ \omega_{ijk} &= \mu_k + \varepsilon_{ijk} \end{aligned} , \tag{8}$$

where $Var(\mu_k) = \sigma_\mu^2$ and $Var(\varepsilon_{ijk}) = \sigma_\varepsilon^2$. The random effects model assumes that:

$$Corr[\omega_{ijk}\omega_{ijl}] = \frac{\sigma_{\mu}^{2}}{(\sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2})}. \quad (9)$$

Consistent and efficient estimates of the parameters in equation (8) are then obtained by performing GLS estimation on the following transformed equation:

$$y_{ijk} - \theta\overline{y}_{ij} = \beta_0(1-\theta) + \beta_1(X_{ijk} - \theta\overline{X}_{ij}) + (\omega_{ijk} - \theta\overline{\omega}_{ij}), \quad (10)$$

where $\theta = 1 - [\frac{\sigma_{\varepsilon}^{2}}{(\sigma_{\varepsilon}^{2} + T\sigma_{\mu}^{2})}]^{1/2}$. (11)

The GLS transformation shown in equation (10) removes the group-wise heteroscedasticity from the sample. By eliminating the heteroscedasticity caused by group-wise dependence in my sample, I achieve unbiased and consistent standard errors.

The theory behind most of the variables is self-explanatory, however, a few require an explanation. *Age65* measures the percentage of those in the neighborhood that are age 65 or older. This type of variable is regularly used as a proxy for a neighborhood's "quietness". The assumption is that consumers are willing to pay more for houses in quiet neighborhoods. *Coast* is a dummy variable equal to 1 if $house_i$ is located in a coastal community. *Pool* and *View* are also dummy variables. In summary, the coefficients on *LivArea*, *Pool*, *View*, *Coast*, *White*, *Age65,* and *Score* are expected to be positive in sign. The coefficients on the remaining variables are expected to have a negative sign.

CHAPTER V

RESULTS

Baseline Specification

Random effects estimates of the coefficients in equation (6) are reported in column 2 of Table 3. Every variable in the baseline specification is statistically significant at the one percent level. The coefficients on all variables except *Age65* are signed as expected.

Because the model has a semi-log functional form, the coefficients can easily be interpreted. Rather than the direct effect on sale price, the coefficients represent the percent change in sale price due to a one-unit increase in the explanatory variable in question. For example, one additional square foot will increase the sale price about 0.03%, all else equal. A coastal location increases the price 21%, and a view increases the price 6.5%, all else equal. A pool increases sale price 8.0%, all else equal.

While percentage changes in price are interesting, knowing the actual dollar change in sale price that amenities cause is more useful. Multiplying the parameter estimates by the sale price will show the change in sale price due to incremental changes in a house's characteristics. Using the mean house price from our sample, $239,097, one additional square foot increases sale price $77, all else equal. Coastal location increases price $50,210, all else equal. The presence of a pool increases price $19,128, all else equal.

The focus of this study is to investigate how school quality affects housing prices. Recall that the measure of school quality used is the district average of 4th and 10th grade reading and math scores on the 1998 Stanford 9 exam. The coefficient on *Score* is statistically significant at the five percent level. The coefficient on *Score* is 0.002; a one-point increase in average test score increases sale price 0.2%, all else equal. At the average price of $239,097, a one-point increase in average test score increases sale price $478, all else equal.

Table 3

Estimation Results Using the Random Effects Estimator

Variable	Baseline Specification Coefficient (z-score)	Expanded Specification Coefficient (z-score)
Structural Characteristics		
Livarea	3.24E-04 (234)	-0.001 (-16.9)
Pool	0.080 (25.9)	0.077 (25.1)
View	0.065 (27.1)	0.065 (27.5)
Neighborhood Characteristics		
Coast	0.21 (58.8)	0.21 (56.3)
White	0.007 (69.0)	0.007 (67.6)
Poverty	-0.008 (-28.1)	-0.008 (-29.4)
Time to work	-0.013 (-16.2)	-0.013 (-15.4)
Age65	-0.003 (-15.1)	0.062 (9.2)
School Characteristics		
Score	0.002 (27.5)	4.80E-05 (0.4)
Interaction Terms		
Score*Livarea	—	1.02E-06 (22.2)
Score*Age65	—	-4.82E-05 (-9.7)
Constant	9.10 (111.2)	11.3 (67.1)

Note: Sample does not include sales of less than $20,000, or sales in which the price per square foot was less than $35/ft^2 or greater than $1,000/ft^2. There are 74,486 observations. R^2 Baseline: 0.69. R^2 Expanded: 0.69. Dependent Variable is ln(Sale Price).

Expanded Specification

Recall from section 3 that in heterogeneous communities, theory suggests that the impact public school quality has on the value of a home depends on the structural and neighborhood characteristics associated with that home. To examine that hypothesis, I expanded the specification given in equation (6) by including a set of interaction terms. Specifically, I added interaction terms measuring the combined effects that school quality and house size (*Score*LivArea*), and school quality and the percent of senior citizens living in a community (*Score*Age65*) have on housing prices. The interaction term (*Score*LivArea*) is designed to capture the differential effect school quality has on large and small homes. As I noted in section 3, theory suggests that large homes should command larger premiums than small homes. Thus, theory predicts that the coefficient on this interaction term should be positive. The interaction term (*Score*Age65*) is designed to capture the fact that households without children (proxied by the percent of the population 65 or older) should be indifferent between homes located in high or low quality school districts. Thus, theory predicts that the coefficient on this interaction term should be negative. That is, the larger the percentage of the population 65 or older within a neighborhood, the smaller the impact of public school quality on housing values.

The results from the expanded specification appear in column 3 of Table 3. With exception to *Score, LivArea*, and *Age65*, the parameter estimates on the variables carried over from equation (5) changed very little. The change in these coefficients is due to their involvement in the interaction terms. *LivArea* changed from 3.24E-04 to –0.001. *Age65* increased from –0.003 to 0.062. The coefficient on (*Score*LivArea*) is 1.02E-06, and the coefficient on (*Score*Age65*) is –4.82E-05.

In the expanded specification, the change in housing price with respect to a change in *Score* is now

$$\frac{\partial \ln(Sale\,Price_{ijk})}{\partial Score_i} = \beta_9 + \beta_{10}(LivArea_{ijk}) + \beta_{11}(Age65_j)\,, \qquad (12)$$

where β_9 is the coefficient on *Score*, β_{10} is the coefficient on (*Score*LivArea*), and β_{11} is the coefficient on (*Score*Age65*). $LivArea_{ijk}$ is the size of $house_i$, and $Age65_j$ is the percent of the population over age 65 in $community_j$. Recall that theory predicts that in heterogeneous communities, school quality should have no direct effect on housing values. In particular, school quality should affect housing values only through the interaction of school quality with the structural and neighborhood characteristics of a home. The results reported in column 3 of Table 3 support that prediction. As anticipated by theory, β_9 is very small and not statistically different from zero. This reinforces the premise that test scores alone should not affect housing prices.

Using the partial derivative of the expanded specification, the effect that a change in *Score* has on sale price, given the mean house size and percent over age 65, is 1.29E-03. That is, a one-point increase in *Score* increases housing prices just over 0.1%, all else equal. Given the mean sale price, a one-point increase in average test score translates to an increase in sale price of $309.

Given equation (12), it is appealing to see how various quantities of *LivArea* and *Age65* affect *Sale Price*. Consider holding house size constant at the mean value of 1,784 ft^2. A one-unit increase in school quality in neighborhoods with 0% age 65 or older increases sale price 0.19%, or $454, all else equal. For neighborhoods with 100% age 65 or older, a one-unit increase in school quality actually *decreases* the sale price of a home by 0.30%

($717), all else equal. When a neighborhood is approximately 39% age 65 or older, school quality has no effect on the sale price of a home.

Now consider holding the neighborhood population age 65 or older constant at the mean (11.9%). If a house is 850 ft^2 in size (a “small” house, for explanatory purposes), a one-unit increase in school quality increases sale price 0.034% ($81), all else equal. By comparison, the sale price of a 3,500 ft^2 house (arbitrarily “large”) increases $717 (0.30%) due to a one-unit increase in school quality, all else equal. These findings are consistent with that proposed by Sonstelie and Portney (1976): changes in the quality of household characteristics affect the price of larger homes more than smaller homes.

Specification Issues

To gauge the sensitivity of the results reported in Table 3, two additional models were estimated. The first model estimates sale price using the number of bedrooms per house in place of a house’s total area. Results are reported in Table 4 of Appendix A. One can argue that, based on the theory presented, number of bedrooms may more accurately represent the demand by households with and without children. Not surprisingly, the results are qualitatively similar to those in Table 3. One additional bedroom increases sale price 19%, or $45,428 (at the mean), all else equal. A one-point increase in average test score increases sale price 0.3%, or $717, all else equal. From the expanded model, a one-point increase in average test score increases sale prices 0.3% ($717), all else equal.

The second model excludes San Diego Unified School District from the sample. This is meant to test the sample’s sensitivity to the inclusion of the largest school district in San Diego County. About 26% of the sample, or 19,724 observations, come from within the San

Diego Unified School District. Estimation results without San Diego Unified School District appear in Appendix B as Table 5. The results are qualitatively similar to those reported in Table 3. Furthermore, by comparing Table 3 with Table 5, we see that the parameter estimates of most explanatory variables are similar in magnitude. However, the magnitude of the parameter estimates for *Score* and *Coast* do change somewhat when San Diego Unified is omitted from the sample. In both the baseline and expanded models, *Score* has a larger effect on sale price when San Diego Unified is removed. *Coast* has a smaller effect on sale price when San Diego Unified is removed.

CHAPTER VI

CONCLUSION

The goal of this study was to measure the capitalization of school quality by housing prices in San Diego County. I used the combined average of fourth and tenth grade math and reading scores on the 1998 Stanford Achievement Test, Ninth Edition as a measure of school quality. The results of my baseline test showed that school quality positively affects housing prices, and this is statistically significant at the one percent level. A one-point increase in average test score increases housing prices by 0.2%, all else equal. To better quantify these results, consider the mean sale price of houses in the sample, which is $239,097. A one-point increase in average test score translates to a $478 increase in sale price, all else equal.

In reality, the affect that school quality has on housing prices acts in a non-linear fashion in conjunction with other housing characteristics. A household with children will choose high school quality over low school quality, so they are willing to pay more for housing near better public schools. However, their location choice is also dependent upon the size of the house. Along the same lines, the greater the senior citizen population is in a neighborhood, the more “peaceful” the neighborhood may be. This, in turn, should increase prices in that neighborhood. But one must ask, “Do senior citizens choose their housing locations based on the quality of local public schools?” One would not think so. So it follows that senior citizens are not willing to pay more for housing near good public schools due to their lower demand for high quality schools.

These thoughts motivated me to expand the specification of my model to measure the combined effect that school quality and the house size, and school quality and the size of the neighborhood senior citizen population, have on housing prices. I found that, given the level of school quality, increasing the house size positively affects housing prices, all else equal. At the mean, a one square foot increase in house size increases sale price by 0.036%, or $86, all else equal. The combined effect of school quality and percent of population over age 65 negatively affects housing prices. Given the mean level of school quality, increasing the percent of population over age 65 decreases housing prices 0.21%, or $504, all else equal.

San Diego County is one of the largest metropolitan areas in the United States. From my results, I can conclude that housing prices definitely capitalize the quality of local public schools. This conclusion is consistent with the results other researchers have found pertaining to other major metropolitan areas throughout the country.

BIBLIOGRAPHY

Allnutt, Donna, and Pennell, Susan. (2000). Crime in the San Diego Region – 2000 Mid-Year Report. SANDAG.

Bergstrom, Theodore C., and Goodman, Richard P. (1973). Private Demands for Public Goods. *American Economic Review, 63*, 280-296.

Bergstrom, Theodore C., Rubinfeld, Daniel L., and Shapiro, Perry. (1982). Micro-Based Estimates of Demand Functions for Local School Expenditures. *Econometrica, 50*, 1183-1206.

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

Brunner, Eric J., Murdoch, James, and Thayer, Mark. (2000). School Finance Reform and Housing Values: Evidence from the Los Angeles Metropolitan Area. *Public Finance and Management*.

Cassel, Eric, and Mendelsohn, Robert. (1985). The Choice of Functional Forms for Hedonic Price Equations: Comment. *Journal of Urban Economics, 18*, 135-142.

Crone, Theodore M. (1998). Housing Prices and the Quality of Public Schools: What Are We Buying? *Business Review of the Federal Reserve Bank of Philadelphia*, 3-14.

Downes, Thomas A. (1992). Evaluating the Impact of School Finance Reform on the Provision of Public Education: The California Case. *National Tax Journal, 45*, 405-419.

Figlio, David N., and Lucas, Maurice E. (2001). What's in a Grade? School Report Cards and House Prices. State of Florida working paper.

Hamilton, Bruce W. (1975). Zoning and Property Taxation in a System of Local Governments. *Urban Studies, 12*, 205-211.

Jud, G. Donald, and Watts, James M. (1981). Schools and Housing Values. *Land Economics, 57*, 459-470.

King, A. Thomas. (1977). Estimating Property Tax Capitalization: A Critical Comment. *The Journal of Political Economy, 85*, 425-431.

Li, Mingche M., and Brown, H. James. (1980). Micro-Neighborhood Externalities and Hedonic Prices. *Land Economics, 56*, 125-141.

Moulton, Brent R. (1986). Random Grouped Effects and the Precision of Regression Estimates. *Journal of Econometrics, 32*, 385-397.

Musgrave, Richard A. (1939). The Voluntary Exchange of Public Economy. *Quarterly Journal of Economics, 52*, 213-217.

Musgrave, Richard A. (December 1955). A Multiple theory of the Budget. A paper presented at the Econometric Society annual meeting.

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

Reinhard, Raymond M. (1981). Estimating Property Tax Capitalization: A Further Comment. *The Journal of Political Economy, 89*, 1251-1260.

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

Samuelson, Paul A. (1954). The Pure Theory of Public Expenditures. *Review of Economics and Statistics, 36*, 387-389.

Samuelson, Paul A. (1955). Diagrammatic Exposition of a Pure Theory of Public Expenditures. *Review of Economics and Statistics, 37*, 350-356.

San Diego Union-Tribune. (July 19, 2000). Special Report: A Measure of Success. Special insert in the *San Diego Union-Tribune*.

San Diego Union-Tribune. (October 5, 2000). Rating the Schools. B4.

Sonstelie, Jon, Brunner, Eric, and Ardon, Ken. (2000). For Better or Worse? School Finance Reform in California. San Francisco: Public Policy Institute of California.

Sonstelie, Jon C., and Portney, Paul R. (1978). Profit Maximizing Communities and the Theory of Local Public Expenditure. *Journal of Urban Economics, 5*, 263-277.

Sonstelie, Jon C., and Portney, Paul R. (1980). Gross Rents and Market Values: Testing the Implications of Tiebout's Hypothesis. *Journal of Urban Economics, 7*, 102-118.

Tiebout, Charles M. (1956). A Pure Theory of Local Expenditures. *The Journal of Political Economy, 64*, 416-424.

APPENDIX A

RANDOM EFFECTS ESTIMATES: BEDROOMS IN PLACE OF LIVAREA

Table 4

Random Effects Estimates: Bedrooms in Place of Livarea

Variable	Baseline Specification Coefficient (z-score)	Expanded Specification Coefficient (z-score)
Structural Characteristics		
Bedrooms	0.19 (117)	-0.27 (-3.4)
Pool	0.17 (45.4)	0.17 (45.3)
View	0.14 (48.8)	0.14 (48.8)
Neighborhood Characteristics		
Coast	0.26 (58.6)	0.25 (57.5)
White	0.009 (72.7)	0.009 (71.3)
Poverty	-0.007 (-19.4)	-0.007 (-19.8)
Time to work	-0.011 (-11.4)	-0.011 (-11.2)
Age65	-0.001 (-4.3)	0.003 (0.3)
School Characteristics		
Score	0.003 (43.4)	0.002 (8.1)
Interaction Terms		
Score*Bedrooms	—	3.44E-04 (5.6)
Score*Age65	—	-2.67E-06 (-0.4)
Constant	9.80 (69.7)	8.41 (25.3)

Note: Sample does not include sales of less than $20,000, or sales in which the price per square foot was less than $35/ft^2 or greater than $1,000/ft^2. There are 74,009 observations. R^2 Baseline: 0.55. R^2 Expanded: 0.55. Dependent Variable is ln(Sale Price).

APPENDIX B

RANDOM EFFECTS ESTIMATES: OMISSION OF SAN DIEGO UNIFIED SCHOOL DISTRICT

Table 5

Random Effects Estimates: Omission of San Diego Unified

School District

Variable	Baseline Specification Coefficient (z-score)	Expanded Specification Coefficient (z-score)
Structural Characteristics		
Livarea	3.00E-04 (205)	-0.002 (-27.8)
Pool	0.097 (28.8)	0.091 (27.0)
View	0.066 (24.9)	0.064 (24.3)
Neighborhood Characteristics		
Coast	0.095 (21.4)	0.085 (19.0)
White	0.003 (17.1)	0.003 (20.5)
Poverty	-0.005 (-12.9)	-0.006 (-13.5)
Time to work	0.004 (3.8)	0.005 (4.9)
Age65	-9.60E-04 (-4.3)	-0.002 (-0.23)
School Characteristics		
Score	0.006 (57.3)	0.002 (13.9)
Interaction Terms		
Score*Livarea	—	1.53E-06 (32.5)
Score*Age65	—	3.83E-07 (0.1)
Constant	3.31 (23.4)	8.27 (37.5)

Note: Sample does not include observations from within the San Diego Unified School District, sales of less than $20,000, or sales in which the price per square foot was less than $35/ft^2$ or greater than $1,000/ft^2$. There are 54,762 observations. R^2 Baseline: 0.70. R^2 Expanded: 0.70.
Dependent Variable is ln(Sale Price).