Geographical Bias in Standardized Testing: Is Cultural and Socioeconomic Bias a Problem of the Past or Are We Simply Not Looking in the Right *Space*?

Manuel S. González Canché*
Higher Education Division
University of Pennsylvania

Author Note

*Mailing address: 208 South 37th Street, Stiteler Hall Room 207, Philadelphia, PA, 19104. Tel. 215-898-0332, email: msgc@upenn.edu

Abstract

Standardized admission tests continue to impact college-going prospects of hundreds of thousands of high school graduates every year. Not only do they affect access to selective institutions but also play a prominent role in merit-aid eligibility. Although these tests were found to be prone to cultural and racial bias a few decades ago due to test wording, the current consensus seems to indicate that this problem has been solved. However, this study hypothesizes that, since spatially-contextualized situations (e.g., poverty, unemployment, education attainment levels in a given place and space) comprehensively affect test-takers' cultural, ethnic, and socioeconomic identities, and these factors are known to affect students' likelihood of perform well in these tests, bias in standardized testing may remain a prevalent yet hidden problem toward social mobility and student diversification in higher education. Two datasets, one at the state-level and one nationally representative, corroborated this hypothesis using spatial analyses that captured the effect of these localized factors. Accordingly, moving forward two implications arise. First, models that use standardized scores to predict students' performance should account for localized factors that have been ignored in the literature to avoid issues of omitted variable bias. The second, and an interrelated issue, is that students' test scores are affected not only by who students are but also by where they experience life. This latter statement indicates that to improve the predictability power of standardized tests, modeling strategies ought to begin accounting for the geolocated factors discussed in this study.

Geographical Bias in Standardized Testing: Is Cultural and Socioeconomic Bias a Problem of the Past or Are We Simply Not Looking in the Right *Space*?

Study Rationale

Despite the recent emergence of test optional and/or test flexible programs, wherein students do not need to, or can decide whether they want to submit their standardized test scores such as SAT/ACT for admission consideration (Belasco, Rosinger, & Hearn, 2014; Syverson, 2007), these standardized admission tests continue to play an important role in college choice, access, and admission decisions. These tests are typically used as inputs for indices that predict students' likelihood of continuation and eventual graduation. In addition, states with generous merit-based grants and scholarships (such as Tennessee and Georgia) use these tests as an important component of eligibility to receive this financial award. Taken together then, these facts indicate that standardized test scores do not only impact access to selective institutions, but also play a prominent role in financial-aid eligibility, hence impacting the college-choice and college-going prospects of hundreds of thousands of high school graduates every year.

Standardized admissions tests originally intended to provide a "*common currency* that allow[ed] admissions officers… to place students on the same footing" (Garvey, 1981, p. 1) when applying to college, regardless of demographic characteristics such as race, ethnicity, or socioeconomic status. Nonetheless, there have been concerns of bias that resulted in "unexplained" differences in standardized test results among test-takers with similar ability levels that only differed in sociodemographic grounds (Clauser & Mazor, 1998; Freedle & Kostin, 1997; Gierl & Khaliq, 2001; Kruse, 2016; Schmitt & Dorans, 1990).These unexplained differences have been commonly referred to as cultural biases (Kruse, 2016) and have been attributed to problems in the design and interpretation of tests (Brown, Reynolds, & Whitaker, 1999; Kruse, 2016).

In the design part, the prevalence of designers of well-educated, middle/upper middle backgrounds arguably led to the development of questions and or wording of tasks that reflected ethnic and racial attributes of similar test-takers' cosmovision compared to those of less privileged backgrounds who have been traditionally referred to as underserved test-takers (*e.g.*, minoritized, first generation, and underrepresented in higher education students). These systematic disconnections in the design-side translated into underserved test-takers being more likely to attain lower scores based on added difficulties in understanding the tasks given the wording of such tasks compared their more privileged counterparts. Accordingly, discrepancies in racial and/or cultural pre-dispositions typically went over and above differences in scholastic aptitudes.

In the interpretation phase, reasoning skill tasks, and or analogies addressed by underserved test-takers yielded lower scores given the evaluators' (and or designers serving as evaluators) cultural and/or racial pre-conceptions or cosmovisions that, once more, had little relation with underserved test-takers scholastic abilities. This negative loop or vicious circle explains that the "concerns of equity and social justice for the students taking these tests have risen" (Kruse, 2016, p. 23) in past decades.

These types of concerns and criticisms motivated research aiming to surpass such potential biases (Freedle, 2003). For example, in 2005 the College Board decided to redesign the SAT to focus more on students' academic preparation than on reasoning skills and removed analogies and quantitative-comparison problems (Epstein, 2009) that could be affected by these cultural differences. Notably, these changes led to today's standing of these tests as being considered robust to these problems, which incidentally also led to their increased weight and presence in admission decisions and merit aid-eligibility, as evidenced by their continued use as "unbiased" independent predictors of scholastic ability, aptitude, and eventual academic success.

This study departs from the existing literature on *bias in testing* in two different yet interrelated aspects. First, it moves beyond issues of cultural, racial and socioeconomic biases at the individual-level by focusing on geographical bias instead. In geographical bias, students' attainment prospects in these standardized scores can be explained by their location, that is by factors measured where students *experience life*, which comprehensively also captures cultural, ethnic/racial and socioeconomic identities. Second, this study moves beyond the traditional use of these tests scores as predictors or as inputs in indices that identify students' probabilities of succeeding in college. Instead, the analyses conducted in this study employ these standardized scores as the outcomes of interest with a twofold purpose. The first is to test whether these indicators are affected by test-takers' geographic location, which constitutes a bias-generating problem (Bivand, Pebesma, & Gómez-Rubio, 2013; Cliff & Ord, 1969, 1972). The second consists of analyzing what localized-level indicators (*e.g.*, poverty-, unemployment-, crime-levels) are the most important predictors of these tests, after correcting for the bias due to test-takers' spatial or geographical dependence along with student-level indicators.

Previous Literature

Previous research suggests that students traditionally underrepresented in higher education students such as first generation in college and students from minoritized racial and ethnic populations (African American, Hispanic, Native American, and Pacific Islander students), and lower income backgrounds, attain lower scores in standardized tests than their majority, higher socioeconomic status peers (Atkinson, 2001; Blau, Moller, & Jones, 2004; Camara & Schmidt, 1999; Crouse & Trusheim, 1988; Fleming, 2002; Freedle, 2003; Geiser & Studley, 2002;

Hoffman & Lowitzki, 2005; Jencks, 1998; Zwick, 2002; Zwick & Green, 2007). This difference in performance has been attributed to access to preparatory courses, private tutors and the resulting increased familiarity with academic vocabulary and the phrasing of questions (Balf, 2014; Kapor & Klein, 2007; Rosner, 2012; Soares, 2012; Zwick, 2004).

Kuncel and Hezlett (2007) discussed the notion of post-graduate education standardized bias in testing. These authors concluded that at this level either there is no evidence of bias based on racial, ethnic, and gender groups. Instead, they concluded that these “tests systematically favor minority groups” (p. 1081). Given the salience of this conclusion, to better interpret the textual quote just provided, it is important to note that, such a form of “advantage” simply reflected that minority groups were predicted to perform better than they did in reality based on their outcomes attained in these tests (Wightman & Muller, 1990). This means that, if one observes two students, one from a minority background and one from a non-minority one, and *both* obtain the same outcome, the non-minority student would be more likely to *succeed* as predicted by these scores, compared to her/his minority counterpart. Considering this finding, then the predictive power of these tests seems to be more relevant for non-minority students, which is a conclusion that is rather different than these tests favoring minority counterparts. While this reduced likelihood in success is not a direct result of test performance, it reflects greater issues in colleges and universities, in particular (e.g., financial aid, micro-aggressions, stereotype threat), and society, in general (e.g., poverty, insecurity, and crime levels), that appear to systematically contribute to minoritized students’ decrease in their prospects for success.

So far, we have two remarkable set of findings with respect to bias in standardized testing. The first is that minoritized and lower income groups tended to underperform in these test scores. This finding is corroborated with test scores associated with the transition from high school to college. The second finding is that if/when minoritized and non-minoritized students happen to attain similar scores, the latter will be more likely to actually succeed as anticipated by those test results.

Based on the relevance of test performance and admission to selective institutions, some scholars have studied whether test preparation is associated with increases in test performance. At this respect, in comprehensive literature reviews authors (Becker, 1990; Leary & Wightman, 983) have concluded that the average gain in students' performance based on coaching (or test) preparation is modest, with about 25% of one standard deviation (Kuncel & Hezlett, 2007). Notably, Kuncel and Hezlett also concluded that test preparation and/or retaking does not affect the predictive power of these tests in program success, with the same caveat found for minoritized students. Kuncel and Hezlett also concluded that more research is needed to better understand factors affecting variation in test scores. They went on to state that "[r]esearchers have examined a number of important outcomes but have not captured other aspects of student performance" (Kuncel & Hezlett, 2007, p. 1081). In line with these authors' conclusions, the extensive literature review conducted for this study indicates that, despite potential differences in race, ethnicity, gender, and socioeconomic status, test performance continues to be predominately conceptualized as inherently individual. That is, race/ethnicity, gender identity and socioeconomic status are all experienced by students. The present study, recognizes and builds from individual level outcomes. However, it also differs from the current literature by arguing that places and spaces where students live capture socioeconomic indicators that complement and affect students' experiences and consequently impact their test performance. As shown in this literature review, the effect of students' geographical location remains an underdiscussed issue that is addressed in this study.

Conceptual Framework

The conceptual framework builds from the notions of concentrated disadvantage (Elijah, 1990; Jargowsky & Tursi, 2015), and geography of opportunity (Pastor, 2001; Tate IV, 2008) or disadvantage (Pacione, 1997). In this framework, participants' (*e.g.*, test-takers) common exposure to spatially-contextualized situations comprehensively affect their cultural, racial, socioeconomic identities along with their greater or lower likelihood of success in life. Using these lenses, students who live in the same geographical area (e.g., neighborhood) are assumed to systemically and systematically be affected by their attendance to the same school districts and schools, exposure to similar levels of localized crime, violence, unemployment, poverty rates, and even access to food options with similar health-quality levels. Hence, it is expected that students' common location and the corresponding shared exposure to "life" not only affects their sociodemographic identities and cultural views but also influence the positive or negative co-variation of their outcomes, namely college-going expectations, safety prospects, and performance in standardized tests scores, to name a few examples.

The application of these frameworks (concentrated disadvantage and geography of opportunity or disadvantage) to test for geographical bias in testing is timely and relevant. If standardized test scores variation do depend on students' location, this finding would be worrisome. These scores are assumed to have become robust to *cultural*, *racial* and socioeconomic *biases* but if they are dependent of students' location, they will be *geographically biased.* If this finding is true, then this form of bias would have been grossly overlooked by the existing literature of bias in testing. From this perspective then, it will follow that based on the concentration of inequality, inequity, poverty, and crime, these standardized scores may have also been actively prone to *racial, cultural,* and, *socioeconomic* biases taking place in spaces overwhelmingly occupied by underserved students who are living and have lived their lives in spaces and places with similar levels of localized resources.

Research Questions.

1. Is there evidence of standardized tests scores co-variation given test-takers' geographical proximity?
    a. If there is evidence of geographical dependence, what subject matter (math, verbal or reading) shows evidence of greater dependence?
    b. Are these results robust or sensitive when analyzing different sources of data, one measured at the state- and the other at a national-level?
    c. If there is evidence of co-variation, was the spatial regression model successful in addressing this problem?
2. What are the most important localized or place-based predictors of variation?
    a. How does the effect of these predictors change when modeling different subject areas and when controlling for participants gender, ethnicity and socioeconomic status?
3. Are the state-level and national-level findings congruent?

Methods

Identification Strategy

The main assumption of the study is that students' location matter in explaining the variation of their standardized test scores. If this is true, then the scores of students who *live in close proximity* would be more likely to have similar scores than students who *live farther away* from one another. Different from more traditional measures of peer effects or network analyses, this co-variation holds regardless of students' friendship status or even in the absence of knowing one another. More specifically, if two students who live one mile apart, for example, have never interacted, they may still have been systematically exposed to the same "standards of living." Accordingly, and following this study's conceptual framework, it may be possible that they attain similar scores in their standardized tests. Following this rationale, as the distance from where they live increases, it is also expected that their outcomes would become less spatially autocorrelated. The mechanism contributing to this variation is based on Tobler's (1970) first law of geography, in which he states that "[e]verything is related to everything else, but near things are more related than distant things" (p. 236).

The identification strategy employed to test for these assumed relationships relies on spatial analyses. Using students' geocoded locations (based on their parents' homes where students were living during high school) several matrices of spatial influence were established. The rules followed to define these matrices of influence are straightforward. If two students live within one mile from each other they would be considered to be "neighbors" as a function of living in spaces that are assumed to share similar levels of resources. In this identification strategy students who live farther away from one mile would not be considered neighbors. Although the one-mile radius identification strategy is assumed to be more rigorous in capturing this outcome dependence, student location density may prove difficult to find neighbors in regions of the country where there are not many test-takers, such as in rural areas. Accordingly, the use of this rule may translate into many students to remain disconnected even if they are indeed exposed to similar experiences. In more densely populated areas, such as cities, however, the one-mile radius would suffice to identify test-takers who live nearby.

In order to address some of the inherent concerns associated with losing participants given their disconnections (as depicted below, the use of spatial analyses would ignore disconnected students), the second identification strategy employed in this study consisted of increasing this radius-based distance to establish neighbors. This second identification strategy expanded the radius to five miles. Regardless of the identification rule employed, it is important to highlight that as distance between participants increase, the expected outcome autocorrelation as a function of such a distance should decay, a notion corroborated in the models presented.

Modeling Strategies

Moran's *I*. The *co-variation of* variables based on participants' location was tested relying on a technique called Moran's *I* (Bivand et al., 2013). As depicted in equation (1) Moran's *I* is calculated as a ratio of the product of the variable of interest (e.g., student's *i* individual test scores or y_i as shown in equation (1)) and the mean of its geographical neighbors (e.g., the scores of the geographical neighbors of student *i*, or y_j as also shown in equation (1)), in both cases with respect to the overall mean ($\bar{y}$) (Bivand et al., 2013). A significant value of *I* yields evidence of more similarity in students' standardized test scores based on their geographical proximity compared to what one would have expected under random chance.

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \quad (1)$$

Moran's *I* requires matrices of influence (w_{ij}), which are build using the radius-based distance approaches discussed above. Moreover, according to Bivand et al. (2013), Moran's *I* is standardized to range from +1 to -1 (Bivand et al., 2013), with positive values indicating that each individual group either systematically performed above (high-performance students clustered with other high-performance students based on their local proximity) or below (low-performance students clustered with other low-performance students also based on their local proximity) with respect to the overall mean ($\bar{y}$). Figures 1(a), 1(b), and 1(c) show students' neighboring structures. Each red dot represents students without proximal connections and students wherein the yellow lines are represented indicate connections at the one- and five-mile radii discussed above, respectively. Figure 1(c) replicates the identification strategy shown in Figure 1(B) but relayed on a nationally representative sample, as discussed below.

Simultaneous Auto-Regressive (SAR). The main purpose of using Moran's *I* is to assess whether there is evidence of students' outcome being more similar given their shared location. If this co-variation is observed, then ordinary least squares models that assumed error terms being identically and independently distributed (iid) would render biased estimates (Bivand, et. al., 2013; Cliff & Ord, 1969, 1972). In spatial analyses a straightforward technique to address for spatial dependence issues is SAR models. Considering that all model specifications corroborated the presence of spatial dependence, SAR modeling was employed to correct for this confirmed spatial co-variation.

The main goal of SAR modeling is to make model residuals to become iid (Bivand, et. al., 2013). Accordingly, these model residuals (ϵ_{i} in equation (2)) are the result of the following modeling procedure:

$$\mathrm{e_i} = \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{w_{ij}}\, \mathrm{e_i} + \epsilon_{\mathrm{i}}, \tag{2}$$

here ϵ_i represents the residuals that are assumed to become independently distributed as a result of the SAR approach. After fitting the SAR models, the residuals may be tested for spatial dependence using a Moran's *I*, which should now be free of co-variation issues if SAR removed this problem. SAR models can incorporate any other regressor or predictor that has is associated with a spatial component. Notably, these spatial components can be measured at different levels, for example, depending on data availability, some indicators can be measures at the zip-code tabulated area, whereas others can be included at the county-level, which coincidentally are the data sources described next.

Data Sources

This study relied on two data sources that account for two standardized tests: ACT and SAT. The first dataset was obtained from 2012 and 2013 SAT test-takers (N=9,066) who enrolled at a Research I university located in the state of Georgia. The second dataset is obtained from the National Center for Education Statistics' (NCES) Education Longitudinal Study (ELS:2002), which documented transitions from high school to college of a nationally representative sample of 10th graders. ELS accounted for ACT scores collected from students' transcripts. If these scores were missing NCES recovered them "through records matching" techniques (ELS, 2014, p. K-4). Before release, NCES made these test scores comparable in Mathematics and Reading to previous datasets, including the National Education Longitudinal Study (1988), High School and Beyond (1980, 2009) to allow analyses of changes across decades.

Both datasets also include students' information that can be georeferenced. The state-level data contain student mailing addresses that enabled the retrieval of their latitude and longitude coordinates with the highest level of precision. In the case of the ELS data, there were no mailing addresses, but these data included the home zip codes of participating students while enrolled in high school. In ELS there were over 14,200 available zip codes, all of which were geo-coded by merging them with the bank of zip codes available from the American Community Survey's (ACS) zip-code shapefile.

The retrieval of this geo-coded information enabled the creation of weight metrices (w_{ij}) discussed above using radii-based approaches (Bivand et al., 2013) to test whether standardized tests showed evidence of being spatially co-dependent given students' proximity.

Both sources of data contained gender, race, and ethnicity indicators, along with a measure of socioeconomic status. In the case of ELS this indicator included an index capturing students' parents' income, job-prestige, and levels of educational attainment. This index ranged from approximately -2 to +2, with negative numbers indicating economic distress. In the case of the state-level dataset the socioeconomic indicator measured both students' *eligibility and reliance* on need-based aid. This indicator was built from information retrieved from the Free Application for Federal Student Aid (FAFSA). This information enabled identification of a subset of students who received need-based aid during their first year of enrollment.

In line with the conceptual framework of this study, the models also included U.S. census-level characteristics that were linked to students' zip-code location. These measures were retrieved from the American Community Survey five-year aggregate estimates across different years. These estimates were selected given their reliability and stability (ACS, 2018) as these data are the result of 60-month data collection period. The ACS data used accounted for the 2007-2011 time period. Models were also fitted with 2005-2009, 2006-2010 rendering consisting results. The selection of 2007-2011 was based on the greatest availability of data points and the assumed influence of these indicators for participant's localized experiences when transitioning from high school to college.

The models also included crime levels at the county-level that were retrieved from the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) Program (2019). This program has the main goal to generate reliable information for use in law enforcement administration, operation, and management. The indicator, at the county-level, was created by dividing all crimes reported at the county-level by the aggregate total of that state. This indicates that, if a given county has a value of 0.20 in this indicator, this means that such a county is accounting for 20% of all the crimes in a given year in that particular county. For consistency purposes this indicator also accounts for five years matching the ACS estimates.

All other indicators shown in Table 1 were measured at the zip-code tabulated area-level. Poverty levels were retrieved from officially reported to the Internal Revenue Service (IRS) under its Individual Income Tax Data. These tax returns were aggregated at the zip-code level by the IRS. From this information, the number of households that reported adjusted gross incomes below the poverty line for a family of four were divided by the total number of households in that zip-code. Accordingly, this coefficient ranges from 0 to 1. Another indicator included in the models is GINI coefficient, which was obtained from ACS and also ranges from 0 to 1, wherein 1 indicates complete income inequality in a given zip-code and zero indicates complete equality. The models also accounted for the proportion of inhabitants in a zip-code who were not U.S. citizens, or who were citizens born in the USA (citizens born abroad is the comparison group). Other indicators obtained from the ACS include Median income, English proficiency (proportion of household that only speak English in a given zip-code), median value of houses, median value of rent, rent to income ratio, proportion of owner-occupied households, proportion of employees who require a college credential, proportion of inhabitants with at least a four-year degree, proportion of households identified as white, and the proportion of households wherein the primary providers were single mothers. Finally, data from the U.S. Department of Agriculture Economic Research Service (2018) was used to measure rurality levels (small town, large town, urban zones, all compared against rural areas obtained from the United States Department of Agriculture). These areas represent approximations that capture the 2000s. The explanation behind the methods used to create these approximations can be found in Ruca Data (2018).

All these indicators were selected from a list of thousands available from the ACS and related sources. Considering that some of them may be highly correlated (that is zones with high crime are likely to also have high poverty levels), in addition to the SAR model specification depicted above, the final set of models relied on a Stepwise Algorithm that detected the best model fit for each outcome as indicated by AIC. All models presented in the study accounted for the indicators shown in Table 1 and then were “trimmed” given this stepwise algorithm (see Venables & Ripley, 2002) as implemented by the MASS package in R. Full models are also available in the appendix.

Findings

As stated above, the analyses relied on the construction of several spatial weight matrices for sensitivity and robustness check purposes. Figure 1 show the 1- and 5-mile radii specifications, with both state- (1(a) and 1(b)) and nation-wide (1(c)) level data. Table 1 contains the summary statistics separated by both, state and nation-wide scopes, and in the case of Georgia, by radii. The first thing to note in the Georgia case is that despite the models including the same initial number of students (9,066), the one-mile radius specification accounted for a lower number of students compared to the five-mile radius (7,285 and 8,713, respectively). This discrepancy is explained by the fact that only students with at least one neighbor were accounted for the inferential analyses. This indicates that as the distance among units is allowed to increase, the resulting number of connections will also increase. This rule is corroborated in Figure 1 and Table 1, where the connections among students located within one-mile from one another accounted for 91,614 connections, whereas the five-mile radius matrix of influence translated into 1.406 million connections. The national sample accounted for 8,230 students and also relied on the five-mile radius approach; however, the resulting total number of connections was

158,192. This lower connectivity is explained by the level of sparsity at the nation-level, as represented in Figure 1(c).

Another value added of including different data sources with differing levels of density is that such a variation between state and nation-level identifications enable empirically testing whether connection density is driving the variation of the results, over and above the capturing of the potential of geographical bias in testing. That is, if the results are robust across density levels and data sources, then the conclusions would be more plausible.

Table 2 shows the results of the Moran's *I* tests. All spatial identification strategies indicated that the standardized test scores are spatially autocorrelated. At the state-level, Table 2 consistently demonstrates that spatial dependence is stronger in the Mathematics than in the Verbal subject area. This systematic dependence is not sensitive to identification rules in terms of significance given the similar probability levels observed across models. Note that in line with the first law of geography, the estimates captured with the one-mile approach are greater in magnitude compared to the ones captured with the five-mile radius. This indicates that as distance increase, dependence based on proximity will tend to decrease.

The nation-level estimates were limited to the 5-mile radius rule because only 2% of the units in the ELS sample lived within a 1-mile radius. Overall, the coefficient magnitudes in ELS were similar to the magnitudes found in the state-level case. In the national sample mathematics scores were also slightly higher spatially dependent in magnitude than those of reading.

Inferential Estimates

Table 3 to 8 contain different specifications across analytic samples. The first two models in each table are unconditional or empty models that were estimated to test whether spatial dependence was addressed when relying on Moran's *I*. The first model is a naïve OLS regression and the second is an empty SAR specification. Note that, consistently, all naïve models presented issues of spatial dependence and all SAR specifications were robust to this problem across different outcomes, scopes, and data sources. This result corroborated the need and importance of using SAR, accordingly, all remaining models included both this SAR estimator along with a test for spatial dependence of the model residuals. This last approach is important as the Lambda indicator across models indicated the presence of spatial dependence issues before the modeling of the residuals, even after including geospatial information in model specification. Once more, all the models successfully addressed this dependence issue as reflected in the insignificant probability values associated with the Moran's *I* coefficients.

Table 9 contains the models resulting from the best fit relying on a stepwise algorithm discussed above (Venables & Ripley, 2002). The inferential results are based on these sets of models. The state-level analyses are included in columns 2 to 5. Across all models, individual level measures were important predictors of outcome variation, which provides evidence of certain groups performing better in these standardized measures than others. This finding is in accordance with past literature. Specifically, the results from this study show that after correcting for error dependence, localized factors, and socioeconomic indicators, White and Asian students performed better in both Mathematics and Verbal standardized tests than the comparison group configured by Hispanic, Native American and African American test-takers. The greater disparities were attained by White students scoring about over 35 points higher than their minoritized counterparts in Verbal. Remarkably, Asian students scored over 72 points higher, on average, than their minoritized counterparts, whereas White students attained scores that were at least 50 points higher. Note that all these students enrolled in a research-intensive institution, which indicates that these disparities may be conservative when comparing test-takers who were not admitted. Another consistent finding was that women attained lower scores than their men counterparts with gaps in Verbal scores that were about five times lower than gap magnitudes observed in Math (-7 points and -37 points, on average, respectively). Finally, in all cases need-based financial aid reliance translated into lower scores, on average, with at least 17 points in Verbal and 21 points in Mathematics. The individual level factors included in the national sample are congruent in all cases except for gender. In the case of the ACT scores, women performed higher in Reading than men, but in Mathematics, once more, women attained lower scores, on average. Similarly, students from wealthier backgrounds attained higher scores in both Reading and Mathematics.

With respect to localized factors, the models consistently indicated that test-takers who live in zip-codes with higher concentrations of inhabitants with bachelor's degrees or more along with those who live in zip-codes with higher median income levels attain better scores in Mathematics, Verbal, and Reading scores. These results hold for both the state and nation-level datasets. Another consistent result across Verbal, Mathematics and Reading outcomes was found for the indicator proportion of citizens who were born in the U.S., compared to citizens who were born abroad (including non-contiguous U.S. territories). In the case of Mathematics, in the state-level, and in the case of both Reading and Mathematics, in the national sample, non-citizens performed poorer than citizens born abroad. With respect to rurality, all students performed better in urban areas across models and tests scores.

All state-level models indicated that test-takers where there is an important presence of single-mother households attain lower scores, on average, with greater reductions in Verbal, reaching up to 60 points reductions. Also, increase in test-takers performance as a function of living in areas where inhabitants predominantly speak English was only achieved in Verbal scores in the state-level analyses.

Notably, poverty levels only reached significance when explaining the variation of Mathematics at the state-level. Students located in zip-codes with high proportion of household living below the poverty line attain lower scores in mathematics, on average. The maximum magnitude this coefficient accounted for a reduction of 49 points (wherein all inhabitants are below this poverty line), with a mean reduction of 19.45 points (i.e., 0.49 * 39.7 pct., see Tables 1 and 9 for these estimates). Similarly, the GINI coefficient for income inequality was an important predictor for reductions in Reading and Mathematics in the nation-wide models, with practically double the magnitude in the latter. Test-takers living in areas with greater concentrations of unemployment consistently attained lower scores in Mathematics at the state level, and in both Reading and Mathematics in the national sample.

Some inconsistent results included increases in the percentage of owner-occupied houses and decreases in SAT scores in Georgia, along with increases in ACT scores with the nation-wide sample. Similarly, SAT (state-level) scores were negatively associated with increases in median home-value in both Mathematics and Verbal, but once more were positively correlated with Mathematics in the national sample (ACT) scores. All the results presented in this section and shown in Table 9 are statistically significant and are robust to spatial dependence as indicated by the Moran's *I* tests of the model residuals.

Limitations

Although all identification strategies intended to capture spatial indicators, a great proportion of such indicators are estimates provided by the ACS, with only IRS and crime FBI data not being estimates but being captured through arguably more robust gathering procedures. This reliance on estimates implies that, most of the results reported herein, while consistent with the conceptual frameworks and previous literature (the latter employing individual-level data) may be conservative. That is, the regression coefficients shown in Table 9 could have made the reported associations to be less relevant in magnitude compared to what students and their families may have been actually experiencing. Although the models incorporated different data sources and distances to establish connections, it is worth noting that the state-level estimates are based on test-takers who were admitted and enrolled at a flagship institution, accordingly, once more the estimates obtained from this sample (a) may only apply to similar test-takers and (b) may once underestimate the effect of spatial dependence. In the case of the nation-wide dataset, its is evident from Figure 1(c) that ELS students were located in predominantly urban areas, accordingly, the same two concerns mentioned for the state-level estimates may also apply. A relevant aspect to be highlighted is that the nationally representative data accounted for test-takers that enrolled at different college and universities (i.e., not research-intensive institutions) yielding congruent results, which may validate the inferences made across models. Finally, the age of the data, especially with the national sample, is a concern and analyses with more recent cohorts are recommended.

Discussion and Implications

The framework of concentrated advantages or disadvantages based on individuals' location (Elijah, 1990; Jargowsky & Tursi, 2015; Pacione, 1997) argues that individuals' prospects for mobility are affected by their places of residence based on the divergent forms of access (or lack of access) to resources associated with experiencing life in particular geolocated contexts. Although this framework is frequently employed in disciplines and fields of study, such as sociology, urban studies, geopolitics, so far it has not been critically utilized to assess the extent to which co-variation in standardized scores occurs as a function of test-takers' location. From this view, methodological perspectives such as spatial analyses employed in conjunction with this conceptual framework of concentrated advantages/disadvantages proved useful to assess and problematize factors directly affecting the reproduction of inequality given participants' location.

Currently, there is no evidence indicating that test optional and test flexible policies are successful in making student bodies more diverse. From this perspective, to the extent that spatial co-variation continues to affect students' performance on these tests, their use in admissions decisions and financial-aid eligibility will continue to fuel social inequality and yield estimates that are biased as a function of ignoring these geographical effects. Failure to consider the effect of location leads to two main forms of bias. The first is omitted variable bias, which by failing to consider important variables, increases the possibility of some of those geographical indicators being associated with some of the observed predictors. The second form of bias is spatial autocorrelation, wherein students' outcomes are dependent simply by living in close proximity to one another. Although this second form of bias is easily addressed using the analytic techniques depicted herein, the use of these approaches remains computationally intensive and expensive as the resulting matrices of influence require computational power that requires bigger processors and more RAM memory. Nonetheless, conceptually, the omission of spatially-related indicators is truly troublesome, for these estimates while imperfect have shown to be significant in explaining the variation of tests scores, even after controlling for students' individual indicators.

In conclusion, given that spatially-contextualized situations comprehensively affect test-takers' cultural, racial, socioeconomic identities along with their greater or lower likelihood of success, bias in standardized testing remains a prevalent problem toward the detriment of social mobility, the strengthening of social reproduction, and the hindering of student diversification in higher education. Accordingly, the main takeaway of this study is that admission and funding decisions that consider these test scores must begin accounting for spatial dependence and should include students' localized indicators. That is, rather than simply stating that these tests have lower predictive power for minoritized populations (Kuncel & Hezlett, 2007; Wightman & Muller, 1990), models should account how other localized factors and spatial dependence affect these predictions, which may indeed be the source of this decrease in predictive power.

References

ACS (2018). When to Use 1-year, 3-year, or 5-year Estimates. *American Community Survey*

Atkinson, R. C. (2001, February 18). Standardized tests and access to American universities. The 2001 Robert H. Atwell Distinguished Lecture, delivered at the 83rd Annual Meeting of the American Council on Education. Washington, DC. Retrieved from: http://www.ucop.edu/news/sat/speech.html.

Balf, T. (2014, March 6). The story behind the SAT overhaul. New York Times Magazine. Retrieved from http://www.nytimes.com/2014/03/09/magazine/the-story-behind-the-sat-overhaul.html.

Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. (2013). *Applied spatial data analysis with r* (Vol. 747248717). Springer.

Belasco, A.S., Rosinger, K.O. & Hearn, J.C. (2014). The test-optional movement at America's selective liberal arts colleges: A boon for equity or something else? *Educational Evaluation and Policy Analysis*, *20*(10), 1-18.

Becker, B. J. (1990). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research*, *60*(3), 373-417.

Blau, J. R., Moller, S., & Jones, L. V. (2004). Why test? Talent loss and enrollment loss. Social Science Research, 33, 409-434

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since Bias in Mental Testing. *School Psychology Quarterly*, *14*(3), 208.

Camara, W. J. & Schmidt, A. E. (1999). Group differences in standardized testing and social stratification (College Board Report No. 1999-5). New York, NY: The College Board.

Cliff, A., & Ord, K. (1969). The problem of spatial autocorrelation. In A. J. Scott (Ed.), *London papers in regional science* (pp. 25–55). London: Pion.

Cliff, A., & Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical analysis*, *4*(3), 267–284.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: issues and practice*, *17*(1), 31-44.

Crouse, J. & Trusheim, D. (1988). The case against the SAT. Chicago, IL: The University of Chicago Press.

Elijah, A. (1990). Streetwise: Race, class, and change in an urban community. *Chicago: University of Chicago*.

ELS (2014). Education Longitudinal Study of 2002 (ELS:2002) Third Follow-Up Data File Documentation: Appendixes. NCES, U.S. DEPARTMENT OF EDUCATION. Available from https://nces.ed.gov/pubs2014/2014364_Appendixes.pdf

Epstein, J. (2009). Behind the SAT-optional movement: Context and controversy. *Journal of College Admission*, 9-19.

Fleming, J. (2002). Who will succeed in college? When the SAT predicts black students' performance. The Review of Higher Education, 25(3), 281-296.

Freedle, R.O. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review*, *73*, 1-43.

Garvey, J.E. (1981). *Testing for college admissions: Trends and issues*. Arlington, VA: Educational Research Service, Inc.

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. Educational Assessment, 8(1), 1-26.

Gierl, M. J., & Khaliq, S. N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests: A confirmatory analysis. *Journal of Educational Measurement*, *38*(2), 164-187.

Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. The Review of Higher Education, 28(4), 455-474.

Jargowsky, P. A., & Tursi, N. O. (2015). Concentrated disadvantage. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences (second edition)* (Second Edition ed., p. 525 - 530). Oxford: Elsevier. Retrieved from https://www.sciencedirect.com/science/article/pii/

Jencks, C. (1998). Racial bias in testing. In C. Jencks & M. Phillips (Eds.), The black-white test score gap. Washington, DC: The Brookings Institution Press, 55-85.

Kapor, M., & Klein, F. K. (2007, November 4). Early admissions policies give children of the rich an edge. The San Francisco Chronicle. Retrieved from http://www.sfgate.com/opinion/article/Early-admissions-policies-give-children-of-the-3234970.php.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science, 315*(5815), 1080-1081.

Kruse, A. J. (2016). Cultural bias in testing: A review of literature and implications for music education. *Update: Applications of Research in Music Education*, *35*(1), 23-31.

Leary, L. F., & Wightman, L. E. (1983). Estimating the relationship between use of test-preparation methods and scores on the Graduate Management Admission Test. *ETS Research Report Series*, *1983*(2), i-61.

Pacione, M. (1997). The geography of educational disadvantage in Glasgow. *Applied Geography*, *17*(3), 169–192.

Pastor, M, Jr. (2001). “Geography and Opportunity.” *America Becoming: Racial Trends and Their Consequences*. Editors Neil J. Smelser, William Julius Wilson, and Faith Mitchell, vol. 1 pp. 435–67. Washington: National Academy Press.

Rosner, J. (2012). The SAT: Quantifying the unfairness behind the bubbles. In J. Soares (Ed.), *SAT wars: The case for test-optional college admissions*. (pp. 104-117). New York, NY: Teachers College Press.

Ruca Data (2018). Ruca Data ZIP Code RUCA Approximation. *US Department of Agriculture*. Retrieved from http://depts.washington.edu/uwruca/ruca-approx.php

Sehmitt, A. P., & Dorans, N. J. (1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement*, *27*(1), 67-81.

Soares, J. A. (2012). For tests that are predictively powerful and without social prejudice. Research and Practice in Assessment, 7, 5-11.

Syverson, S. (2007). The role of standardized tests in college admissions: Test-optional admissions. *New Directions for Student Services*, 118, 55-70.

Tate IV, W. F. (2008). “Geography of opportunity”: Poverty, place, and educational outcomes. *Educational Researcher*, *37*(7), 397-411.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*(sup1), 234-240.

Uniform Crime Reporting (UCR) Program. (2019). Federal Bureau of Investigation. See https://www.fbi.gov/services/cjis/ucr

U.S. Department of Agriculture Economic Research Service. (2003). Codes, rural-urban continuum. *US Department of Agriculture*. Retrieved from http://depts.washington.edu/uwruca/ruca-codes.php

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S.* Fourth edition. Springer.

Wightman, L. F., & Muller, D. G. (1990). An Analysis of Differential Validity and Differential Prediction for Black, Mexican American, Hispanic, and White Law School Students. LSAC Research Report Series.

Zwick, R. (2002). Fair game? The use of standardized admissions tests in higher education. New York, NY: RoutledgeFalmer.

Zwick, R. (2004). Is the SAT a "Wealth Test"? The link between educational achievement and socioeconomic status. In R. Zwick (Ed.), Rethinking the SAT: The future of standardized testing in university admissions (pp. 203-216). New York, NY: RoutledgeFalmer.

Zwick, R., & Green, J. G. (2007). New perspectives on the correlation of scholastic assessment test scores, high school grades, and socioeconomic factors. Journal of Educational Measurement, 44(1), 1-23.

Table 1: Descriptive statistics of predictor and control variables

Variable	n	x̄	s	Min	Max
One-Mile indicators, Georgian test-takers					
SATM	7285	615.6	82.9	260.0	780.0
SATV	7285	540.3	81.2	250.0	740.0
Pct_belowPov	7285	38.9	9.1	18.5	100.0
gini	7285	0.4	0.1	0.3	0.6
notcitizen	7285	0.1	0.1	0.0	0.4
citizenbornusa	7285	0.9	0.1	0.4	1.0
onlyEnglish	7285	0.8	0.1	0.2	1.0
medhouseval	7285	239.8	115.7	44.9	932.1
medrent	7285	969.1	185.6	422.0	2001.0
grossrentincome	7285	0.2	0.1	0.0	0.4
owneroccup	7285	0.7	0.1	0.0	1.0
nonprofrelated	7285	2.4	0.5	0.0	5.0
motheronlyhousehold	7285	0.2	0.1	0.0	0.8
smalltown	7285	0.0	0.2	0.0	1.0
largetown	7285	0.0	0.2	0.0	1.0
urban	7285	0.9	0.3	0.0	1.0
WhiteNH	7285	0.7	0.2	0.0	1.0
BachOrMore00s	7285	0.4	0.2	0.0	1.0
MediIncome	7285	34.5	10.0	0.0	77.7
PctUnemployed	7285	0.1	0.0	0.0	0.2
PctCrimePop	7285	0.0	0.1	0.0	1.0
RACE.D.W	7285	0.8	0.4	0.0	1.0
RACE.D.A	7285	0.1	0.3	0.0	1.0
SEX.F	7285	0.6	0.5	0.0	1.0
needaideligible	7285	0.3	0.4	0.0	1.0
Five-Mile indicators, Georgian test-takers					
SATM	8713	611.6	83.3	260.0	780.0
SATV	8713	537.7	81.6	200.0	740.0
Pct_belowPov	8713	39.7	9.4	17.4	100.0
gini	8713	0.4	0.1	0.3	0.6
notcitizen	8713	0.1	0.1	0.0	0.4
citizenbornusa	8713	0.9	0.1	0.4	1.0
onlyEnglish	8713	0.8	0.1	0.2	1.0
medhouseval	8713	231.4	120.2	34.2	1000.0
medrent	8713	947.4	211.5	344.0	2001.0
grossrentincome	8713	0.2	0.1	0.0	0.5
owneroccup	8713	0.7	0.1	0.0	1.0
nonprofrelated	8713	2.5	0.5	0.0	6.4
motheronlyhousehold	8713	0.2	0.1	0.0	0.8
smalltown	8713	0.0	0.2	0.0	1.0
largetown	8713	0.1	0.2	0.0	1.0
urban	8713	0.9	0.3	0.0	1.0
WhiteNH	8713	0.7	0.2	0.0	1.0
BachOrMore00s	8713	0.4	0.2	0.0	1.0
MediIncome	8713	33.3	10.4	0.0	83.2
PctUnemployed	8713	0.1	0.0	0.0	0.2
PctCrimePop	8713	0.0	0.1	0.0	1.0
RACE.D.W	8713	0.8	0.4	0.0	1.0
RACE.D.A	8713	0.1	0.3	0.0	1.0
SEX.F	8713	0.6	0.5	0.0	1.0
needaideligible	8713	0.3	0.5	0.0	1.0
Five-Mile indicators, ELS test-takers					
MathScoreNELS	8238	46.3	13.7	14.7	79.3
ReadScoreNELS	8238	30.3	9.4	9.8	50.6

Comparison groups: *citizens not born in the USA, **Rural
***In $ thousands

Continued on next page...

... table 1 continued

Variable	**n**	$\bar{\mathbf{x}}$	**s**	**Min**	**Max**
Pct_belowPov	8238	43.1	10.3	16.1	100.0
gini	8238	0.4	0.1	0.1	0.7
notcitizen	8238	0.1	0.1	0.0	0.5
citizenbornusa	8238	0.8	0.1	0.2	1.0
onlyEnglish	8238	0.7	0.2	0.0	1.0
medhouseval	8238	265.8	176.5	10.0	1000.0
medrent	8238	983.7	302.3	235.0	2001.0
grossrentincome	8238	0.2	0.1	0.0	0.7
owneroccup	8238	0.7	0.2	-0.1	1.0
nonprofrelated	8238	2.8	0.6	0.0	6.4
motheronlyhousehold	8238	0.3	0.2	-0.3	1.1
smalltown	8238	0.0	0.1	0.0	1.0
largetown	8238	0.0	0.2	0.0	1.0
urban	8238	1.0	0.2	0.0	1.0
WhiteNH	8238	0.7	0.2	0.0	1.0
BachOrMore00s	8238	0.3	0.2	0.0	1.0
MediIncome	8238	29.4	9.6	0.0	101.2
PctUnemployed	8238	0.1	0.0	0.0	0.5
PctCrimePop	8238	0.1	0.2	0.0	1.0
SESF1	8238	0.1	0.8	-2.1	2.0
White	8238	0.5	0.5	0.0	1.0
AsianNatHawPcf	8238	0.1	0.3	0.0	1.0
Female	8238	0.5	0.5	0.0	1.0

Comparison groups: *citizens not born in the USA, **Rural

Table 2: Moran's I test results for neighboring structures

	I	E(I)	var(I)	St. Dev.	p-val.
Geographical Network Dependence State-level					
1-Mile Radius SAT Math*	0.156	-0.0001	6.81E-05	19.013	$< .001$
5-Mile Radius SAT Math**	0.130	-0.0001	1.52E-05	33.389	$< .001$
1-Mile Radius SAT Verbal*	0.083	-0.0001	6.81E-05	10.051	$< .001$
5-Mile Radius SAT Verbal**	0.071	-0.0001	1.52E-05	18.221	$< .001$
Geographical Network Dependence National Sample					
5-Mile Radius Math Test***	0.169	-0.0001	3.89E-05	27.06	$< .001$
5-Mile Radius Reading Test***	0.164	-0.0001	3.89E-05	26.318	$< .001$

*7,299 test-takers with 91,614 connections,
**8,713 test-takers with 1,405,158 connections,
***8,238 test takers with 158,192 connections

Table 3: One-Mile Radius Models Verbal

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	540.31***	540.31***	578.19***	686.63***	629.58***	589.24***
	(0.95)	(0.95)	(4.15)	(61.51)	(65.57)	(67.75)
Pct_belowPov			−0.97***	−0.38*	−0.15	0.02
			(0.10)	(0.18)	(0.19)	(0.20)
Gini				107.10**	63.00	79.20*
				(34.23)	(35.98)	(37.02)
notcitizen				−149.59	−127.65	−114.73
				(78.68)	(81.05)	(81.29)
citizenbornusa				−235.55***	−237.95**	−214.37**
				(68.98)	(78.13)	(78.87)
onlyEnglish				124.19*	112.41*	100.50
				(53.41)	(55.15)	(55.39)
medhouseval				0.00	−0.00	−0.00*
				(0.00)	(0.00)	(0.00)
medrent				−0.00	0.00	0.00
				(0.01)	(0.01)	(0.01)
grossrentincome				−3.26	−1.51	14.18
				(18.54)	(18.52)	(19.97)
owneroccup				−61.71***	−45.66**	−52.52**
				(16.73)	(17.15)	(17.35)
nonprofrelated				−2.46	3.09	3.18
				(3.01)	(3.36)	(3.37)
motheronlyhousehold				−114.54***	−55.97**	−54.71*
				(15.93)	(21.44)	(21.59)
smalltown				−3.42	−4.40	−4.05
				(12.93)	(12.93)	(12.94)
largetown				3.03	1.80	1.50
				(12.41)	(12.40)	(12.41)
urban				7.94	5.36	6.22
				(11.89)	(11.95)	(11.95)
WhiteNH					34.85***	41.32***
					(10.31)	(11.55)
BachOrMore					44.81**	30.19
					(15.47)	(16.74)
MediIncome						0.00**
						(0.00)
PctUnemployed						7.41
						(80.58)
PctCrimePop						17.53
						(20.51)
R^2	0.00	0.00	0.01	0.04	0.04	0.04
Adj. R^2	0.00	0.00	0.01	0.03	0.04	0.04
Num. obs.	7285	7285	7285	7285	7285	7285
RMSE	81.24	81.24	80.76	79.82	79.72	79.70
AIC	NA	84661	84599	36070	36142	35961
Log Likelihood	NA	-42327.29	-42295.44	-18003.23	-18038.95	-17950.35
Num. obs.	7285	7285	7285	7285	7285	7285
Lambda	NA	0.146***	0.126***	-0.830***	-2.141***	-0.573**
Moran's I	0.083***	-0.003	-0.004	-0.003	-0.001	-0.0005
Moran's Std.Dev	10.049	-0.635	-0.481	-0.399	-0.448	-0.448

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: One-Mile Radius Models Mathematics

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	615.61***	615.61***	670.29***	1053.43***	984.46***	957.79***
	(0.97)	(0.97)	(4.21)	(61.32)	(65.28)	(67.31)
Pct_belowPov			−1.41***	−0.16	0.15	0.38
			(0.11)	(0.18)	(0.19)	(0.20)
Gini				130.88***	71.72*	71.22
				(34.12)	(35.81)	(36.78)
notcitizen				−480.85***	−461.61***	−422.92***
				(78.43)	(80.68)	(80.77)
citizenbornusa				−478.53***	−504.41***	−448.34***
				(68.76)	(77.78)	(78.36)
onlyEnglish				58.99	50.74	21.69
				(53.24)	(54.91)	(55.03)
medhouseval				0.00	−0.00	−0.00**
				(0.00)	(0.00)	(0.00)
medrent				−0.00	0.01	−0.00
				(0.01)	(0.01)	(0.01)
grossrentincome				−63.99***	−61.29***	−22.40
				(18.48)	(18.44)	(19.84)
owneroccup				−53.54**	−31.62	−36.26*
				(16.68)	(17.07)	(17.24)
nonprofrelated				−8.61**	−1.26	−2.24
				(3.00)	(3.34)	(3.35)
motheronlyhousehold				−121.62***	−35.52	−23.22
				(15.88)	(21.35)	(21.46)
smalltown				7.13	6.10	3.85
				(12.89)	(12.87)	(12.85)
largetown				13.81	12.32	10.15
				(12.37)	(12.35)	(12.33)
urban				10.41	7.48	9.93
				(11.85)	(11.89)	(11.88)
WhiteNH					53.25***	43.17***
					(10.26)	(11.48)
BachOrMore					56.14***	19.19
					(15.40)	(16.64)
MediIncome						0.00***
						(0.00)
PctUnemployed						−343.78***
						(80.06)
PctCrimePop						22.86
						(20.38)
R^2	0.00	0.00	0.02	0.08	0.09	0.09
Adj. R^2	0.00	0.00	0.02	0.08	0.08	0.09
Num. obs.	7285	7285	7285	7285	7285	7285
RMSE	82.90	82.90	81.91	79.57	79.36	79.18
AIC	36086	36100	36084	36070	36142	35961
Log Likelihood	-18010.76	-18017.81	-18009.87	-18003.23	-18038.95	-17950.35
Num. obs.	1780	1780	1780	1780	1780	1781
Lambda	NA	-0.572***	-0.716***	-0.830***	-2.141***	-0.573**
Moran's I	0.158***	-0.007	-0.006	-0.003	-0.001	-0.0005
Moran's Std.Dev	19.175	-0.501	-0.597	-0.399	-0.448	-0.448

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Five-Mile Radius Models Verbal

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	537.69^{***}	537.69^{***}	582.21^{***}	667.89^{***}	592.78^{***}	559.15^{***}
	(0.87)	(0.87)	(3.77)	(52.38)	(55.55)	(56.78)
Pct_belowPov			-1.12^{***}	-0.47^{**}	-0.14	0.03
			(0.09)	(0.17)	(0.18)	(0.19)
Gini				108.00^{***}	56.22	75.65^{*}
				(28.55)	(30.10)	(30.94)
notcitizen				-103.23	-69.86	-68.78
				(67.56)	(68.80)	(68.88)
citizenbornusa				-251.44^{***}	-216.47^{**}	-195.17^{**}
				(61.18)	(68.91)	(69.20)
onlyEnglish				161.42^{***}	132.96^{**}	116.65^{*}
				(46.35)	(48.34)	(48.53)
medhouseval				-0.00	-0.00^{*}	-0.00^{***}
				(0.00)	(0.00)	(0.00)
medrent				-0.00	-0.00	-0.00
				(0.01)	(0.01)	(0.01)
grossrentincome				-6.17	-4.86	12.63
				(15.42)	(15.41)	(16.53)
owneroccup				-57.21^{***}	-41.73^{**}	-50.78^{***}
				(13.38)	(13.65)	(13.92)
nonprofrelated				-3.23	2.70	2.73
				(2.52)	(2.80)	(2.81)
motheronlyhousehold				-111.20^{***}	-60.81^{***}	-60.33^{***}
				(13.55)	(17.61)	(17.67)
smalltown				-1.73	-2.58	-2.68
				(8.63)	(8.63)	(8.63)
largetown				6.62	5.46	4.74
				(8.36)	(8.35)	(8.35)
urban				12.26	9.35	9.90
				(7.95)	(8.01)	(8.03)
WhiteNH					30.91^{***}	37.00^{***}
					(8.79)	(9.62)
BachOrMore					57.83^{***}	37.48^{*}
					(13.27)	(14.82)
MediIncome						0.00^{***}
						(0.00)
PctUnemployed						-6.35
						(67.49)
PctCrimePop						19.12
						(18.27)
R^2	0.00	0.00	0.02	0.04	0.04	0.05
Adj. R^2	0.00	0.00	0.02	0.04	0.04	0.04
Num. obs.	8713	8713	8713	8713	8713	8713
RMSE	81.58	81.58	80.90	79.94	79.81	79.77
AIC	NA	84661	84599	36070	36142	35961
Log Likelihood	NA	-42327.29	-42295.44	-18003.23	-18038.95	-17950.35
Num. obs.	7285	7285	7285	7285	7285	7285
Lambda	NA	0.146^{***}	0.126^{***}	-0.830^{***}	-2.141^{***}	-0.573^{**}
Moran's I	0.072^{***}	-0.003	-0.004	-0.003	-0.001	-0.0005
Moran's Std.Dev	33.389	-0.635	-0.481	-0.399	-0.448	-0.448

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: Five-Mile Radius Models Mathematics

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	611.58^{***}	611.58^{***}	674.26^{***}	952.93^{***}	852.12^{***}	830.40^{***}
	(0.89)	(0.89)	(3.82)	(52.35)	(55.41)	(56.55)
Pct_belowPov			-1.58^{***}	-0.30	0.16	0.38^{*}
			(0.09)	(0.17)	(0.18)	(0.18)
Gini				117.65^{***}	44.49	56.87
				(28.53)	(30.03)	(30.82)
notcitizen				-307.33^{***}	-266.23^{***}	-249.69^{***}
				(67.52)	(68.63)	(68.60)
citizenbornusa				-418.54^{***}	-386.10^{***}	-349.13^{***}
				(61.14)	(68.73)	(68.91)
onlyEnglish				97.94^{*}	64.39	40.13
				(46.32)	(48.22)	(48.33)
medhouseval				0.00	-0.00^{*}	-0.00^{***}
				(0.00)	(0.00)	(0.00)
medrent				-0.01	-0.00	-0.01
				(0.01)	(0.01)	(0.01)
grossrentincome				-57.65^{***}	-55.48^{***}	-21.67
				(15.41)	(15.37)	(16.46)
owneroccup				-34.20^{*}	-12.12	-20.82
				(13.37)	(13.62)	(13.87)
nonprofrelated				-9.79^{***}	-1.58	-2.51
				(2.52)	(2.79)	(2.80)
motheronlyhousehold				-103.93^{***}	-27.55	-20.83
				(13.54)	(17.56)	(17.60)
smalltown				7.64	6.71	6.20
				(8.62)	(8.61)	(8.59)
largetown				19.94^{*}	18.47^{*}	17.41^{*}
				(8.35)	(8.33)	(8.32)
urban				16.13^{*}	12.46	15.42
				(7.94)	(7.99)	(7.99)
WhiteNH					48.41^{***}	43.39^{***}
					(8.76)	(9.58)
BachOrMore					78.63^{***}	38.04^{**}
					(13.24)	(14.76)
MediIncome						0.00^{***}
						(0.00)
PctUnemployed						-266.13^{***}
						(67.22)
PctCrimePop						22.23
						(18.20)
R^2	0.00	0.00	0.03	0.08	0.09	0.09
Adj. R^2	0.00	0.00	0.03	0.08	0.09	0.09
Num. obs.	8713	8713	8713	8713	8713	8713
RMSE	83.32	83.32	82.00	79.89	79.61	79.44
AIC	36086	36100	36084	36070	36142	35961
Log Likelihood	-18010.76	-18017.81	-18009.87	-18003.23	-18038.95	-17950.35
Num. obs.	1780	1780	1780	1780	1780	1781
Lambda		-0.572^{***}	-0.716^{***}	-0.830^{***}	-2.141^{***}	-0.573^{**}
Moran's I	0.130^{***}	-0.007	-0.006	-0.003	-0.001	-0.0005
Moran's Std.Dev	18.221	-0.501	-0.597	-0.399	-0.448	-0.448

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7: Five-Mile Radius Models Verbal, National Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(Intercept)	30.25***	30.25***	40.85***	31.94***	27.78***	32.07***	35.44***
	(0.10)	(0.10)	(0.43)	(2.88)	(2.92)	(3.01)	(2.81)
pro_belowPov			−0.25***	−0.08***	0.00	−0.04	−0.02
			(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
gini				17.51***	1.18	−1.57	−5.04
				(2.60)	(3.03)	(3.10)	(2.89)
notcitizen				−11.38***	−12.77***	−9.93**	−8.64**
				(3.37)	(3.36)	(3.42)	(3.19)
citizenbornusa				−10.58***	−11.93***	−10.69***	−8.25***
				(2.39)	(2.42)	(2.46)	(2.29)
onlyEnglish				7.19***	5.92***	6.31***	3.43**
				(1.36)	(1.38)	(1.39)	(1.29)
medhouseval				−0.00*	−0.00*	−0.00	0.00
				(0.00)	(0.00)	(0.00)	(0.00)
medrent				−0.00	−0.00	−0.00	−0.00
				(0.00)	(0.00)	(0.00)	(0.00)
grossrentincome				−1.80	0.43	−0.32	−0.73
				(1.51)	(1.52)	(1.59)	(1.48)
owneroccup				2.29*	3.05**	4.62***	2.77**
				(1.05)	(1.04)	(1.10)	(1.02)
nonprofrelated				−0.77***	0.21	0.02	0.05
				(0.22)	(0.24)	(0.25)	(0.23)
motheronlyhousehold				−10.54***	−4.43**	−2.36	−1.78
				(1.25)	(1.50)	(1.54)	(1.43)
smalltown				0.28	0.79	1.25	0.93
				(1.78)	(1.77)	(1.76)	(1.64)
largetown				3.18**	2.60*	2.67*	2.43*
				(1.19)	(1.18)	(1.18)	(1.10)
urban				3.73***	2.70**	3.27**	2.64**
				(1.03)	(1.03)	(1.03)	(0.96)
WhiteNH					3.17***	2.78***	2.70***
					(0.65)	(0.66)	(0.62)
BachOrMore00s					12.25***	14.40***	7.00***
					(1.26)	(1.51)	(1.42)
MediIncome						−0.15***	−0.14***
						(0.03)	(0.03)
PctUnemployed						−17.51***	−13.98**
						(5.18)	(4.82)
PctCrimePop						0.05	0.14
						(0.59)	(0.55)
SESF1							4.76***
							(0.13)
R^2	0.00	0.00	0.07	0.11	0.12	0.12	0.24
Adj. R^2	0.00	0.00	0.07	0.11	0.12	0.12	0.24
Num. obs.	8230	8230	8230	8230	8230	8230	8230
RMSE	9.41	9.41	9.07	8.91	8.85	8.83	8.22
Lambda		-0823***	-0.616***	-0.530***	-.541***	-0.673**	
Moran's I	0.165***	-0.003	-0.007	-0.005	-0.001	-0.0009	
Moran's Std.Dev	23.311	-0.418	-0.378	-0.579	-0.493	-0.534	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8: Five-Mile Radius Models Mathematics, National Sample

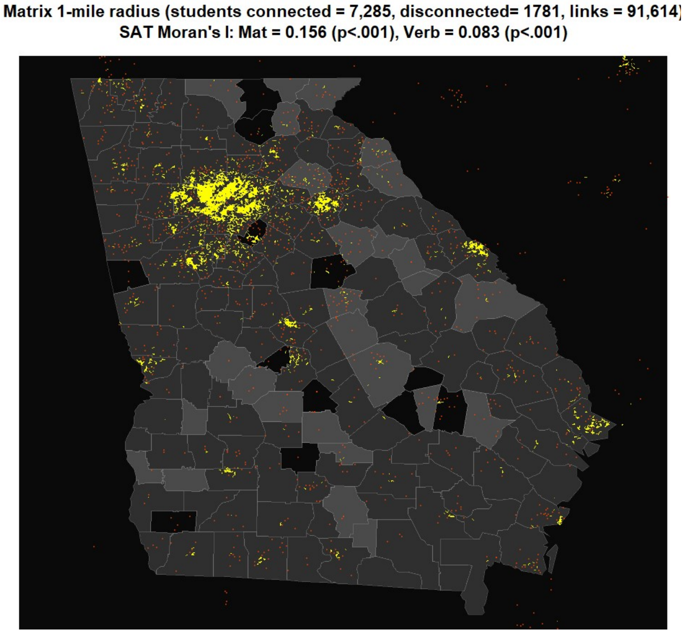
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(Intercept)	46.28***	46.28***	63.03***	58.38***	51.77***	58.61***	63.16***
	(0.15)	(0.15)	(0.62)	(4.14)	(4.19)	(4.32)	(4.06)
pro_belowPov			−0.39***	−0.06*	0.07*	−0.02	0.01
			(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
gini				21.08***	−4.53	−9.06*	−13.72**
				(3.74)	(4.35)	(4.45)	(4.18)
notcitizen				−27.48***	−29.64***	−23.97***	−22.24***
				(4.84)	(4.82)	(4.91)	(4.61)
citizenbornusa				−22.22***	−24.41***	−21.66***	−18.38***
				(3.43)	(3.48)	(3.52)	(3.31)
onlyEnglish				10.12***	8.16***	8.55***	4.68*
				(1.96)	(1.99)	(1.99)	(1.87)
medhouseval				0.00	0.00	0.00	0.00**
				(0.00)	(0.00)	(0.00)	(0.00)
medrent				−0.00	−0.00**	−0.00	−0.00
				(0.00)	(0.00)	(0.00)	(0.00)
grossrentincome				−0.13	3.37	2.12	1.55
				(2.18)	(2.18)	(2.29)	(2.15)
owneroccup				2.45	3.64*	6.62***	4.13**
				(1.51)	(1.50)	(1.57)	(1.48)
nonprofrelated				−1.26***	0.27	0.01	0.05
				(0.32)	(0.35)	(0.35)	(0.33)
motheronlyhousehold				−21.28***	−11.57***	−8.11***	−7.33***
				(1.80)	(2.15)	(2.20)	(2.07)
smalltown				0.09	0.89	1.68	1.25
				(2.55)	(2.53)	(2.53)	(2.38)
largetown				1.84	0.93	1.04	0.72
				(1.71)	(1.70)	(1.70)	(1.59)
urban				3.77*	2.15	3.19*	2.34
				(1.48)	(1.48)	(1.48)	(1.39)
WhiteNH					5.07***	4.21***	4.11***
					(0.93)	(0.95)	(0.89)
BachOrMore00s					19.17***	23.20***	13.24***
					(1.81)	(2.17)	(2.06)
MediIncome						−0.26***	−0.25***
						(0.04)	(0.04)
PctUnemployed						−26.93***	−22.17**
						(7.43)	(6.98)
PctCrimePop						−1.21	−1.08
						(0.84)	(0.79)
SESF1							6.41***
							(0.19)
R^2	0.00	0.00	0.09	0.12	0.14	0.14	0.24
Adj. R^2	0.00	0.00	0.09	0.12	0.14	0.14	0.24
Num. obs.	8230	8230	8230	8230	8230	8230	8230
RMSE	13.67	13.67	13.07	12.81	12.70	12.66	11.90
Lambda		-0.572***	-0.816***	-0.730***	-.741***	-0.673**	
Moran's I	0.169***	-0.004	-0.005	-0.004	-0.002	-0.0003	
Moran's Std.Dev	29.067	-0.351	-0.475	-0.239	-0.578	-0.548	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

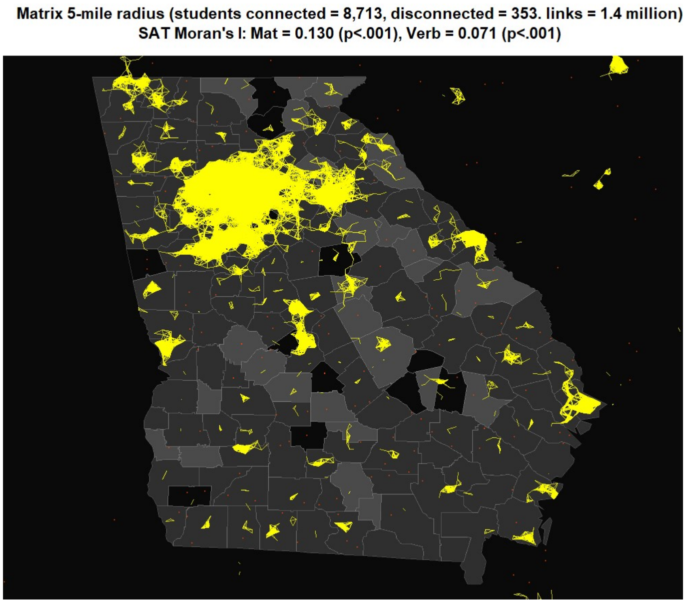
Table 9: Best Model Fit

	1-m SATV	1-m SATM	5-ms SATV	5-ms SATM	5-ms ELSV	5-ms ELSM
(Intercept)	552.78***	870.96***	553.06***	732.38***	31.58***	51.59***
	(18.87)	(53.94)	(20.79)	(45.81)	(2.19)	(3.11)
citizenbornusa	−142.31*	−310.99***	−131.20**	−195.13***	−4.67**	−9.08***
	(56.38)	(44.97)	(55.02)	(38.23)	(1.50)	(2.51)
onlyEnglish	122.37**		114.05**			
	(44.38)		(41.83)			
medhouseval	−0.04**	−0.03	−0.04**	−0.06***		0.002***
	(0.01)	(0.02)	(0.01)	(0.01)		(0.001)
grossrentincome	28.47		28.46			
	(14.77)		(17.51)			
owneroccup	−51.80***	−35.15*	−48.98***	−22.40	2.89***	5.15***
	(13.25)	(14.30)	(11.62)	(12.20)	(0.79)	(1.20)
motheronlyhousehold	−59.51***	−33.15*	−59.58***	−30.74*		
	(12.62)	(16.01)	(14.90)	(13.32)		
urban	8.65**		8.43*	10.74**	2.95**	1.67*
	(3.54)		(3.92)	(4.03)	(0.78)	(0.71)
BachOrMore00s	40.83***	49.12***	41.51***	57.76***	7.32***	12.79***
	(11.13)	(12.31)	(12.63)	(10.98)	(1.24)	(1.77)
MediIncome	0.53*	0.87***	0.46*	0.95***	0.12***	0.23***
	(0.22)	(0.24)	(0.25)	(0.21)	(0.02)	(0.03)
RACE.D.W	37.41***	50.06***	37.95***	52.68***		
	(2.70)	(2.83)	(2.94)	(2.60)		
RACE.D.A	10.17***	72.63***	9.72*	73.29***		
	(1.93)	(3.98)	(4.16)	(3.77)		
SEX.F	−7.07***	−38.05***	−7.38***	−36.33***		
	(1.72)	(1.78)	(1.88)	(1.64)		
needaideligible	−17.32***	−21.05***	−17.48***	−21.42***		
	(1.95)	(2.07)	(2.18)	(1.86)		
Pct_belowPov		-0.42*		−0.49**		
		(0.18)		(0.13)		
notcitizen		−339.23***		−178.95**	−6.05***	−11.91*
		(68.35)		(57.77)	(1.73)	(4.11)
PctUnemployed		−104.90**		−119.21*	−15.86***	−21.78***
		(45.80)		(48.21)	(4.15)	(6.33)
smalltown			-6.88			
			(5.78)		2.82**	
largetown				7.70		
				(5.64)	(0.95)	
PctCrimePop				27.18		
				(16.95)		
gini					−6.29**	−11.91***
					(2.38)	(3.55)
SESF1					4.30***	5.74***
					(0.13)	(0.19)
White					4.21***	6.39***
					(0.22)	(0.33)
AsianNatHawPcf					1.82***	7.68***
					(0.29)	(0.42)
Female					1.24***	−1.49***
					(0.18)	(0.25)
WhiteNH						2.46**
						(0.73)
R^2	0.09	0.20	0.08	0.20	0.27	0.29
Adj. R^2	0.08	0.20	0.08	0.20	0.27	0.29
Num. obs.	7285	7285	8713	8713	8238	8238
RMSE	78.07	74.06	77.99	74.41	8.05	11.53
AIC	84153	83378	100650	99777	57701	63637
Log Likelihood	-42060.63	-41673.03	-50307.32	-49869.38	-28834.35	-31801.4
Lambda	0.059***	0.09***	0.15***	0.22***	0.148***	0.125***
Moran's I	0.0004	0.003	0.002	-0.003	0.0004	0.003
Moran's Std. Dev	-0.09	-0.29	-0.70	-0.76	-0.70	-0.51

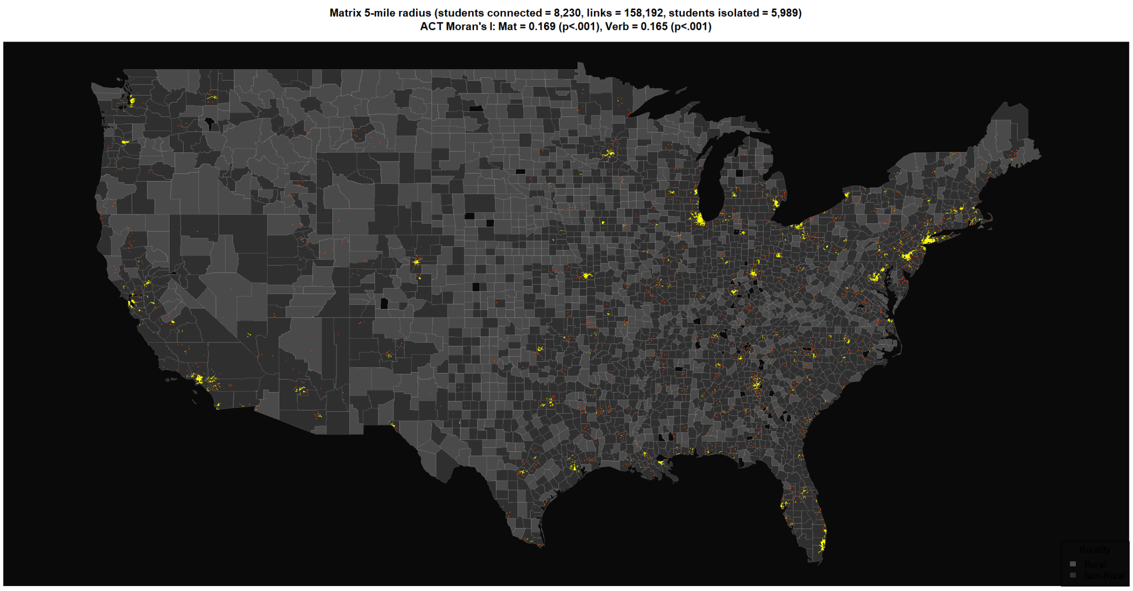
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$



(a) Georgian test-takers 1-mile radius



(b) Georgian test-takers 5-mile radius



(c) ELS test-takers 5-mile radius

Figure 1: Geographical Network Matrices Analyzed Using Moran's I and SAR models