Journal of Urban Mathematics Education
December 2020, Vol. 13, No. 2, pp. 26–41

EDITORIAL

Using Large-Scale Datasets to Amplify Equitable Learning in Urban Mathematics

Eduardo Mosqueda
University of California Santa Cruz

Saúl I. Maldonado
San Diego State University

The United States has administered national mathematics assessments since 1973, and since then researchers have identified a relationship between students' achievement, racial-ethnic background, and enrollment in advanced mathematics courses (Carpenter et al., 1983). Patterns of performance on achievement measures such as the National Assessment of Educational Progress (NAEP) influence policy decisions, and disaggregating data by students' socioeconomic status, race-ethnicity, and English-language competence remains a federal priority (de Brey et al., 2019). Although policymakers have used disaggregated student assessment results to justify funding decisions (Trujillo, 2016), researchers have argued that the persistent disparities in achievement trends by disaggregated groups reproduces educational inequalities (Gutiérrez, 2007; Johnson, 2002).

Disrupting the mathematics achievement disparities of students minoritized by socioeconomic status, race-ethnicity, and English-language competence requires an explicit consideration of schools' designation as either urban, suburban, or rural. In the United States, schools in metropolitan communities comprised of more than 50,000 persons often serve high percentages of students minoritized by socioeconomic status, race-ethnicity, and English-language competence (Lippman et al., 1996). An association exists between mathematics achievement and students' income, racial-ethnic, and language background; however, student characteristics cannot completely explain differential mathematics achievement, and consideration of urban school contexts is required (Capraro et al., 2013).

To challenge deficit thinking often associated with urban schools and to provide a tool for researchers studying urban education, Milner (2012) provided a three-category typology for classifying urban schools: urban intensive, urban emergent, and urban characteristic. Urban intensive describes schools in large metropolitan cities, urban emergent describes schools in cities with populations under 1 million people, and urban characteristic describes schools not located in cities but facing

EDUARDO MOSQUEDA is Associate Professor in the Department of Education at the University of California Santa Cruz, 1156 High Street, Santa Cruz, California, 95064; email: mosqueda@ucsc.edu. His research interests include the mathematics education of emergent bilingual learners and equity in urban school contexts.

SAÚL I. MALDONADO is an Assistant Professor in the Department of Dual Language and English Learner Education at San Diego State University, 5500 Campanile Drive, San Diego, California 92182; email: smaldonado@sdsu.edu. Maldonado investigates the intersection of language, literacy, K–12 achievement and evaluation.

contextual challenges similar to those of urban intensive and urban emergent schools (e.g., limited instructional resources, inadequate teacher preparation, and increases in the population of linguistically minoritized (LM) students). Milner's typology of urban schools has recently been amplified to include a) population/location/geography, b) enrollment, c) demographic composition of students, d) resources in schools, e) disparities and educational inequality, and f) social and economic context (Welsh & Swain, 2020). In addition, Welsh and Swain (2020) have also found that LMs comprise 11% to 15% of students enrolled across all three categories of urban schools. The urban context, therefore, provides important information regarding the learning environment in which students operate.

From our perspective, it is critical that researchers distinguish how students' individual characteristics, schools' urban contexts, and the provision of advanced mathematics courses influence mathematics achievement trends. The opportunities-to-learn (OTL) conceptual framework informs our research of mathematics achievement in urban schools. Researchers have described OTL as equitable access to the structural conditions that develop all students' learning (Stevens & Grymes, 1993; Tate, 2001). We consider OTL an appropriate framework for our additive perspective of students in urban schools negotiating mathematics learning opportunities as mechanisms of power (Martin et al., 2010).

The United States Department of Education has an agency dedicated to providing education data and research to the public: The Institute of Education Sciences (IES). Established in 2002, IES replaced the Office of Educational Research and Improvement to address the "mismatch between what education decision makers want from the education research and what the education research community is providing" (Whitehurst, 2003, p. 13). IES administers programs such as the What Works Clearinghouse as well as four centers: a) The National Center for Education Research, b) The National Center for Education Statistics, c) The National Center for Education Evaluation and Regional Assistance, and d) The National Center for Special Education Research. In this article, we present suggestive guidance for accessing and using large-scale datasets from the National Center for Education Statistics (NCES) to examine secondary mathematics achievement in U.S. urban schools.

Our discussion's focus is on using large-scale datasets produced by NCES to examine mathematics achievement, but we intend for this paper to be a guide for researchers in both accessing and using large-scale data in general for the purpose of developing a quantitative research agenda that examines students' mathematics achievement in urban secondary school contexts in the United States. Our purpose is to provide methodological considerations and analytical suggestions for researchers using NCES datasets to examine mathematics outcomes of students minoritized by race-ethnicity, socioeconomic status, and English-language background. We are specifically interested in sharing suggestive guidance to equity-informed researchers committed to urban mathematics education issues. In what follows, we provide a description of the NCES secondary longitudinal studies and guidance for accessing this data. We then offer methodological considerations for analyses with NCES datasets and conclude with analytical suggestions for studies focused on urban school contexts. For clarity, a data source is represented in bold italics, and variables are all upper case.

NCES Secondary Longitudinal Studies

Researchers interested in questions about secondary students' mathematics achievement in urban schools may benefit from accessing data from the following longitudinal surveys: a) National Longitudinal Study of the High School Class of 1972 [***NLS-72***], b) High School and Beyond [***HS&B***], c) National Education Longitudinal Study of 1988 [***NELS:88***], d) Educational Longitudinal Study of 2002 [***ELS:2002***], and e) High School Longitudinal Study of 2009 [***HSLS:09***]. We display a historical overview of the research design of the five longitudinal studies along with information about data collection years and students' corresponding ages and years in school in Figure 1.

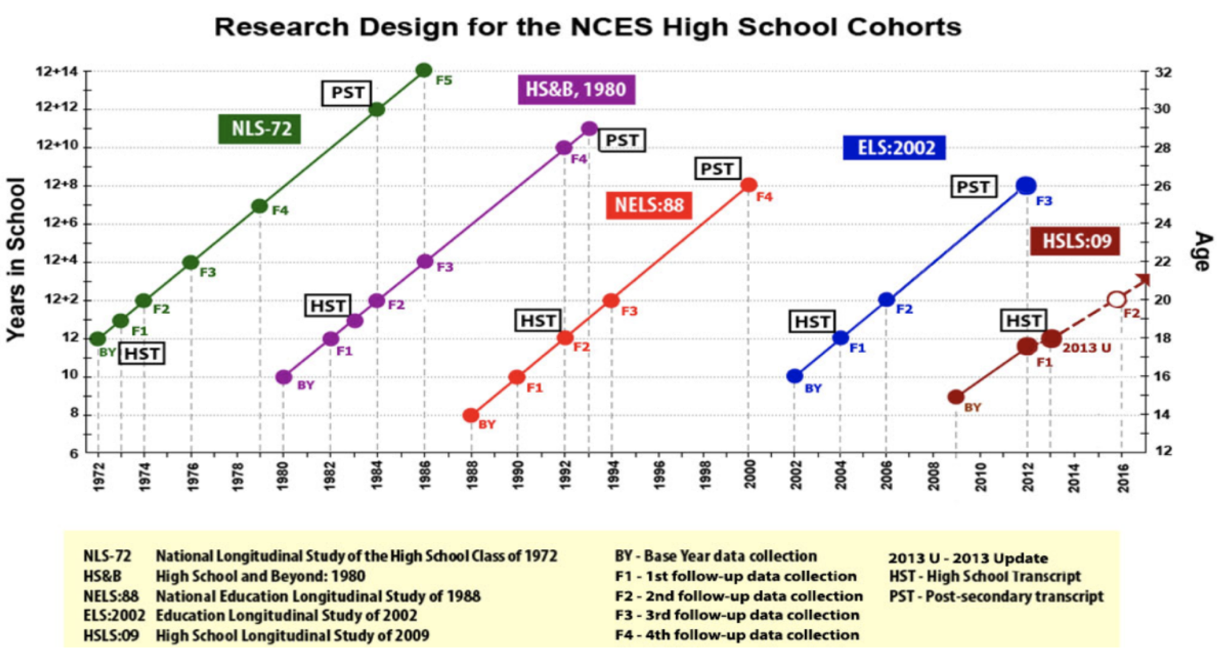

Figure 1. Research Design for NCES Secondary Longitudinal Studies

Beginning with the ***NLS-72***, NCES secondary longitudinal studies provide researchers with almost fifty years of quantitative information to analyze students' mathematics achievement. Data collection has concluded for ***NLS-72***, ***HS&B***, and ***NELS:88***, but data collection is ongoing for ***ELS:2002*** and ***HSLS:09***. The most recent NCES secondary dataset, the Middle Grades Longitudinal Study of 2017–18 [***MGLS:2017***], is currently collecting mathematics assessment data of students, beginning in the sixth grade and concluding in the eighth grade.

Accessing and Using NCES Datasets

The IES website offers the public free unrestricted access to data, publications, products, and data tools associated with NCES. Using the online codebook data tool, researchers may download variables of interest in school-level or student-level data as well as create syntax files. IES offers free Distance Learning Dataset Training modules online to teach researchers how to acquire, access, and explore NCES datasets as well as how to conduct analyses with specific data tools and statistical software. We recommend researchers start with the Common Modules, such as "Analyzing NCES Complex Survey Data," and continue with the 34-minute module "Introduction to the NCES Longitudinal Studies: 1972–2020." After understanding the objectives of the various secondary longitudinal studies, we suggest researchers further familiarize themselves with one of the NCES datasets.

Prior to downloading any public-use data for analysis, we encourage researchers to review the data file documents that describe specific details about each dataset, such as appropriate design weights and information about the suppression of all direct and indirect identifiers that may compromise the confidentiality of participating persons and schools. Additionally, we recommend researchers review tables previously created by NCES and published online prior to downloading the variable lists for longitudinal studies of interest to identify if research questions will require access to restricted-use school- or student-level data. Researchers from organizations in the United States, such as universities and government agencies, that are interested in analyzing restricted-use data must participate in a comprehensive application process for a license, which includes signed documents, such as notarized affidavits of nondisclosure, as well as the completion of an online training course. Students require a faculty advisor to submit the application as the principle project officer, who will work in partnership with both a senior official from the university's sponsored projects office as well as a systems security officer from the university's technology department to finalize the application. After reading the restricted-use data procedures manual, researchers submit the license application with signed original documents to the IES Data Security Office. Each license application permits up to seven users of the IES data, including students, provided all users access and analyze the data in the same location. If the restricted-use data license is approved, the requested data file(s) are mailed to the license holder as an encrypted CD-ROM/DVD for exclusive use on a secured standalone desktop computer that is not connected to a network or modem, and the data must remain in a locked office that is only accessible to persons listed on the license. Additionally, IES requires that researchers analyzing restricted-use NCES data round unweighted sample size numbers and submit all reports, tables, or presentations for potential disclosure review prior to dissemination.

In our research, we have analyzed secondary students' achievement data from ***NELS:88***, ***ELS:2002***, and ***HSLS:09***. When we used ***HSLS:09*** data to examine secondary mathematics achievement in U.S. urban schools, the student racial-ethnic background composite variable [X1RACE] and the mathematics standardized theta score [X1TXMTSCOR] were available on the public-use files, but the student native language background composite variable [X1NATIVELANG] was only accessible on the restricted-use files. Using the public-use data, we ran descriptive statistics, such as frequencies, percentages, crosstabs, and correlation tables, to determine if the number of ***HSLS:09*** cases would be appropriate for our research questions.

Methodological Considerations

Complex Sampling

The NCES typically uses complex sample designs in its data collection approach that include the following three strategies: stratification, clustering, and a multi-stage approach. For example, in describing the sampling approach used in ***HSLS:09***, Ingels and colleagues (2011) note the following:

> In the base-year survey, students were sampled through a two-stage process. First, stratified random sampling and school recruitment resulted in the identification of 1,889 eligible schools. A total of 944 of these schools participated in the study, resulting in a 55.5 percent (weighted) or 50.0 percent unweighted response rate. In the second stage of sampling, students were randomly sampled from school ninth-grade enrollment lists, with 25,206 eligible selections (or about 27 students per school). (p. v)

The NCES notes that their use of a complex sampling design is to increase the efficiency in measuring specific subsamples in a population. For example, stratification is used to ensure that different subgroups are adequately represented in the sample. Stratification involves dividing a sampling frame into relevant subgroups prior to the sample selection (Schneider et al., 2007). Although complex sampling strategies are useful to ensure sufficient numbers of underrepresented observations in a sample, such sampling strategies also give more weight to particular observations that are disproportionately included relative to their representation in the overall population. Therefore, the weights that are included in each dataset must be included in an analysis to adjust for any oversampling (Thomas et al., 2005). The NCES uses sampling weights to indicate the relative contribution of each observation in order to produce adequate population-level estimates. For example, if a student in a dataset is assigned a weight of 1050, this means that the student represents 1,050 students in the population who have the characteristics used in the sampling design, such as racial-ethnic background and grade level. The NCES provides the sampling design information needed to use complex variance estimation software to compute estimates of variance that reflect the complex sample design of the data collection process. If this information is not used, or if used incorrectly, the results of hypothesis testing, or the *p* values, will also be incorrect.

Sample weights are used to account for sample selection processes, meaning they adjust for the fact that not all units had an equal probability of selection into the sample. Sample weights can be adjusted for the fact that nonresponse may be greater among certain subgroups of the population. This adjustment is important because when there are differential patterns of nonresponse the data can be biased, or not representative of either the population or subgroups of interest. If all sources of nonresponse bias are accurately captured in the nonresponse weight adjustments, the end result produces estimates that represent the target population.

Data Clustering

Clustering of observations results from the selection of groups of units, such as first selecting schools then selecting students. If clusters of students are internally homogeneous, meaning students within schools are more similar than students across schools, then the estimates of the overall variance on measures will be lower than is the case if a simple random sampling strategy was employed (Muthén & Satorra, 1995; Thomas et al., 2005). Multi-stage cluster sampling can also be used to facilitate multilevel analysis of the relationships between distinct levels of data. Multilevel modeling is a variance estimation technique that is appropriate for complex sample designs, as it addresses the clustered samples into the analytical models (Muthén & Satorra, 1995; Raudenbush & Bryk, 2002). For each participant in the sample, the total score on a dependent variable is decomposed into an individual (or within-group) component and a between-group component. The decomposition of variables from the sample data into their component parts can be used to compute a within-group covariance matrix (i.e., the covariance matrix of the individual deviations from the group means) and a between-group covariance matrix (i.e., the covariance matrix of the disaggregated group means), and the variation at each level can then be explained simultaneously with sets of predictors at each level of the data structure (Raudenbush & Bryk, 2002; Thomas & Heck, 2001). There are multiple statistical software packages, such as HLM, MLwinN, SPSS, STATA, and R, to name a few, that can take advantage of the nesting of students (level 1) within schools (level 2), which will adjust for the variance estimation appropriate for complex designs used by NCES.

Causal Inference

The American Educational Research Association has availed a report that provides researchers (and funding agencies) with a set of guidelines for evaluating various methods and analytic approaches for estimating causal effects from large-scale observational datasets that can inform policy (Schneider et al., 2007). These guidelines focus on the following analytic approaches for estimating causal effects: fixed-effects models, instrumental variables, propensity scores, and regression discontinuity (e.g., Schneider et al., 2007). All of these methods provide useful strategies for eliminating bias in observational designs, such as the secondary longitudinal studies produced by NCES.

In research designs that have a clear treatment variable, one popular method for addressing selection bias in such observational datasets is propensity score matching (PSM). Although participants in a specific treatment activity are not randomly assigned to specific interventions in observational data sets, students may differ systematically in such assignments. This results in selection bias in the estimation of the treatment effects. PSM analysis can be used to correct for such selectivity bias (Rosenbaum & Rubin, 1984; Singer & Willett, 2003). The benefit of PSM is that it produces a conditional probability, or propensity, of being in the treatment or control group based on a set of observed variables (Rosenbaum & Rubin, 1984). An additional benefit to PSM is that it can be utilized to match students based on key observable characteristics in order to minimize selection bias in the estimated treatment effects and simultaneously support causal inferences based on the results for participants in the treatment relative to those in the comparison group. An important limitation of using propensity scores to address selection bias in observational datasets is that PSM only accounts for observed covariates and does not account for unobserved characteristics of participants (Murnane & Willett, 2011; Rubin, 1997). Despite the method's limitations, we believe PSM improves researchers' ability to more effectively compare treatment and comparison group students in large-scale dataset observational studies.

Analytical Suggestions from Our Published Studies

Analyzing Multiple Levels of Data Simultaneously

There are multiple measures that reflect the urban school context at the individual level and at the school level that influence mathematics achievement, often times with negative effects on the overall outcomes. In other words, characteristics of minoritized students, including race/ethnicity, language status, and socioeconomic status (SES), can all depress achievement performance on standardized tests. For instance, an NCES study shows that in 2015 the average mathematics score on the

NAEP assessment for White 12^{th}-grade students was 22 points higher than the test performance of their Latinx peers and 30 points higher than their Black peers (McFarland et al., 2019). This NAEP study also showed that the mean mathematics score was 115 for 12^{th}-grade English language learner (ELL) students, reflecting a negative difference of 38 points, on average, relative to the mean score of 153 for their non-ELL counterparts. In addition, schools with higher concentrations of low-SES students had lower aggregated test scores in mathematics. The NAEP study showed that, on average, the aggregated mathematics scores for 12th-grade students in high-poverty schools was 129, a mean score that is markedly lower than the average scores for 12th-grade students in mid-high-poverty schools (145), mid-low-poverty schools (154), and low-poverty schools (164) (McFarland et al., 2019). The disparities highlighted in this NAEP report across race and ethnicity, student family incomes of varying levels, and school poverty levels and between ELLs and non-ELLs suggest that urban mathematics education analyses must account for income at the individual level and also at the school level by including measures of low-income student concentration.

The inclusion of socioeconomic measures at the individual level in addition to the socioeconomic concentration and makeup of a school's student population, which is also an important marker of a school's urbanicity, is consistent with extant research that has found SES exerted the largest influence on academic achievement rather than schools' racial-ethnic composition (Rumberger & Palardy, 2005). Other research by Orfield and Lee (2007) has focused on the connection between urban school contexts and academic achievement as influenced by within-school factors, such as students' racial-ethnic segregation and family income. These results showed that students who attended schools in which the mean of students' SES was high received increased academic benefits, while students in schools where the mean SES was low showed lower achievement performance outcomes.

The effects of the concentration of ELLs in specific schools has been identified as an additional factor that negatively moderates achievement. Gándara and Orfield (2010) researched how disproportionate segregation of Latinx ELLs negatively impacted achievement in addition to negatively influencing students' social and emotional development. Another study on the segregation of Latinx ELLs showed their concentration in schools was associated with lower achievement outcomes in mean scores and was negatively influenced in terms of content knowledge, English-language fluency, academic-language fluency, and literacy development (Gifford & Valdés, 2006). Results from these studies show the importance of including race/ethnicity, language status, and SES school-level variables in analyses of achievement in urban schools.

Most NCES datasets include multiple urbanicity-related measures for studies to account for individual-background and school-context variation that may negatively influence academic performance outcomes. These predictors may also help account for selectivity bias in the sample. Such measures include individual-level race/ethnicity, gender, and SES. At the school level, we suggest researchers use aggregate urbanicity measures, which include the following: whether the school is public or private, the percentage of low-income students within each school, the number of students that qualify for free or reduced lunch (used as a proxy for poverty), the percentage of ELLs, and the percentage of teachers that are not fully credentialed. In addition, the NCES datasets provide a robust measure of family SES that includes income, parents' occupational status, and educational attainment. For these reasons, we recommend the use of the SES indicator over the unidimensional family income variable in studies of urban schools.

In a study titled "Systematized discrimination: The relationship between students' linguistic minority status, race-ethnicity, opportunities to learn, and college preparatory mathematics," Mosqueda and colleagues (under review) investigated differential patterns of opportunities to learn among secondary school students minoritized by race/ethnicity, SES, and LM status in urban schools. The researchers accounted for individual level covariates that included SES, gender, race-ethnicity, LM status, and each student's self-reported level of English-language proficiency (ELP). In addition, the following urban school context measures were integrated into the analysis: school type indicator (1 = Public and 0 = Catholic or other private), urban designation (0 = Suburban or Rural and 1 = Urban), SCHOOLSES (Percent of 12th grade students eligible for free or reduced lunch), percent LEP (Percent of LM students in 12th grade), percent college track (Percent of 12th grade students in the college/academic track), and percent credentialed teachers (Percent of 12th grade credentialed teachers). Each of these measures was highlighted in the literature reviewed related to the mathematics achievement of ethnically and linguistically minoritized students in urban schools.

Variability in English-Language Proficiency

Our work has examined the mathematics achievement patterns of LM students in urban schools. We have primarily focused on two indicators of linguistically minoritized status: whether students are native English speakers (coded as NON-ELL or as Linguistic Minority status) and, when available, we have concurrently examined the effect of students' self-reported level of ELP. Mosqueda (2010) as well as Mosqueda and Maldonado (2013) used ***ELS: 2002*** data to examine the relationship between mathematics achievement and academic track placement or course-taking patterns of Latinx students in 12th grade, respectively. In Mosqueda (2010), the results showed that the effect of academic track placement and mathematics achievement differed as a function of both LM status and each LM student's degree of ELP, relative to non-LM students. The findings suggested that Latinx ELLs with lower ELP benefitted less from high-track placement than both Latinx ELLs with high degrees of ELP and their non-ELL peers. Mosqueda and Maldonado (2013) found that

LM status and ELP were strongly related to mathematics course-taking, and students who took higher level mathematics courses benefitted more than those in lower level math courses. The findings from both studies suggest that the variability in the range of ELP from a low to a high degree is a better indicator of the effects of language than a binary indicator of LM status. Perhaps because LM students are included with students of emergent levels of proficiency and others within this same reference category, LM students may have ELP that is near English-proficient and can potentially compare to the English proficiency of a native English speaker.

There are multiple language background measures in NCES datasets that can inform studies of mathematics in urban schools (Mosqueda & Maldonado, 2013). For example, LM status can be coded from the variable F1STLANG, a survey question that asked students, "Is English your native language (the first language you learned to speak when you were a child)?" (Ingels et al., 2004). Additionally, we have used students' self-reported ELP measure in both ***NELS:88*** and ***ELS:2002***. In these datasets, four ordinal dimensions included how well students "understand spoken English," "speak English," "read English," and "write English" and comprise ELP. For each of these dimensions of English proficiency, students provided one of following ordinal responses: "Very well," "Well," "Not well," or "Not at all" (Ingels et al., 2004). In order to only account for the variation in ELP of LM students, we have differentiated among the level of English proficiency of LM students, using the cross-product of LM_{ij} * $ENGPROF_{ij}$. The variable $ENGPROF_{ij}$ was a weighted composite resulting from principal components analysis of the four dimensions of ELP in the survey. We also acknowledged a limitation of using such self-reported measures; however, we note that the reliability of similar measures has been established in large-scale studies of immigrant students (Portes & Rumbaut, 2001). Because there are no measures of ELP at a national level, this method is one of the most useful measures of ELP available. Although ***NELS:88*** and ***ELS:2002*** datasets both include an ELP indicator for LM students, ***HSLS:09*** does not include a measure of ELP. Maldonado and colleagues (under review) used PSM to examine the relationship between Latinx students' English-language and immigrant background, teachers' mathematical proficiency practices, and urban school contexts in ***HSLS:09 data***. Researchers also used three restricted-use variables to determine students' English-language background: NATIVELANG (indicates the language the student first learned to speak), DUALLANG (indicates if student is multilingual), and EBL (indicates if student is currently an emergent bilingual ELL and enrolled in an ELL Program). Findings from the quasi-experimental multi-level linear regression models showed that the relationship between teachers' mathematical practices, urban school contexts, and Latinxs' mathematics achievement was not influenced by English-language background. Without measures that account for the variability of students' ELP, researchers may neglect an important dimension in urban students' mathematics achievement.

Immigrant Generational Status

When NCES studies like ***HSLS:09*** do not include ELP measures for LM students, researchers may consider analyzing the interrelated nature of ELP and immigrant generational status. We have concurrently analyzed the mediating effects of English-language background and immigrant generational status on students' mathematics achievement using both ***ELS:2002*** and ***HSLS:09*** data. When we analyzed ***ELS:2002*** data (a study that includes ELP variability for LMs), the effect of immigrant generational status on mathematics achievement was not statistically significant (Mosqueda & Maldonado, 2013). When we analyzed ***HSLS:09*** data (a study that does not include ELP variability for LMs), the effect of immigrant generational status on mathematics achievement was statistically significant (Maldonado et al., under review). Although closer examinations of the degree of confoundedness between immigrant generational status and English-language background were important, our results imply that immigrant generational status information is also an important factor to consider when analyzing urban students' mathematics achievement, particularly for NCES studies that do not include ELP information.

Maldonado and colleagues (under review) used three restricted-use ***HSLS:09*** variables to determine students' immigrant background: SCOUNTRY (country in which student was born), P1COUNTRY (country in which Parent1 was born), and P2COUNTRY (country in which Parent2 was born). Using SCOUNTRY, P1COUNTRY, and P2COUNTRY data, we created three dichotomous immigrant generational status variables: FIRST, SECOND, and THIRD. FIRST indicates students who were born in a country other than the United States and had at least one parent who was born in a country other than the United States. SECOND indicates students who were born in the United States and had at least one parent who was born in a country other than the United States. THIRD indicates cases wherein students and both parents were born in the United States. Our coding scheme for students' immigrant generational status was consistent in our analyses of both ***HSLS:09*** data and ***ELS:2002*** data.

Another reason to consider the inclusion of immigrant generational status information in studies of urban school students' mathematics achievement is the potential effect of unobserved values connected to interrelated student background characteristics of race-ethnicity, English-language proficiency, and immigrant generational status. Prior research has found that academic achievement and immigrant aspirational optimism provide explanations for how the second generation outperforms first and third generation immigrant students (Kao & Tienda, 1995; Portes & Rumbaut, 2001). Results from these studies show the importance of including individual-level variables of language background and immigrant generational background in analyses of achievement in urban schools.

Inadequately Defined Indicators

There will be cases in which specific variables of interest to a study may not be adequately defined and so are unable to provide useful results to an analysis. For example, when analyzing ***ELS: 2002*** data, we were interested in a variable that would gauge whether or not the presence of a mathematics teacher with specialized Limited English Proficiency training was associated with higher Latinx ELLs' mathematics test scores. Our analysis revealed that the presence of teachers with specialized Limited English Proficiency training did not have a statistically significant effect on the assessment outcomes of Latinx ELLs (Mosqueda, 2011). However, this finding was not conclusive, because the variable for the language training of teachers was inadequately defined. In the ***ELS:2002*** survey, the associated question specifically asked whether "teachers had at least 8 hours of specialized training over the last 3 years" in working with ELLs. For those teachers reporting having had such training, some may have attended a single-day (eight-hour) workshop for training on teaching ELLs over the last three-year period. Alternatively, it could mean that a teacher earned a graduate degree in second-language instruction. Given the wide range of this measure, more work is needed in future NCES studies to create more reliable measures of specialized professional development for teachers. Our more recent studies attempt to go beyond existing information about mathematics teachers' certification, credentialing, years of experience, and participation in professional development. For example, we are now analyzing teachers' specific pedagogical practices that are consistent with the strands of mathematical proficiency (Maldonado et al., under review).

Illuminating Inequity to Improve Urban Mathematics Achievement

Informed by OTL, this article offers analytical guidance for education researchers interested in examining the structural conditions that influence mathematics learning and achievement in urban schools. Specifically, we argue for researchers to consider how the intersectional interplay of students' SES, race-ethnicity, and English-language background and urban school contexts influences mathematics achievement. Our purpose in this article was to provide researchers examples of prior analyses to inform methodological considerations when using NCES datasets to analyze secondary students' mathematics achievement in urban schools.

Our rationale for our recommendations is to emphasize the importance of capturing the effects of specific characteristics in urban school contexts that influence students' mathematics achievement. Second, we consider it important to inform researchers that studies that include school-level percentages of students minoritized by SES, race-ethnicity, and English-language background in statistical models but omit examinations of curricular resources or pedagogical practices may reproduce a deficit discourse of urban communities (Capraro et al., 2013; Moses & Cobb, 2001). Policy and practice inequities in urban schools are complex, and we recognize that students' participation in advanced mathematics courses is only one variable in a multitude of social, cultural, and political factors that influence mathematics teaching and learning. We consider our research aligned with studies that promote equitable access and meaningful participation in mathematics learning as a civil right (Moses & Cobb, 2001; Perry et al., 2003).

We are hopeful that researchers interested in students' mathematics achievement in urban schools will consider accessing and using NCES studies to deepen our collective understanding beyond IES summary statistics featured in reports such as NAEP, TIMMS, and tables from *The Condition of Education*. We believe that studies of large-scale data that are attentive to the methodological considerations of complex sampling, data clustering, and causal inference will contribute nuanced perspectives of promising policy and practice directions for improving urban students' mathematics achievement. Additionally, we believe this article details specific data sources and student, school, and community variables that are important to consider when analyzing achievement in urban schools.

Although we have exclusively analyzed standardized mathematics achievement scale scores, NCES datasets offer multiple measures of students' achievement, including grade point averages and transcripts with course-taking information. We recommend researchers consider the appropriateness of all outcome variables in relation to research questions as well as methodological fit and analytical design. For example, the advantage of using a standardized mathematics achievement scale score may offer readers useful analytical interpretations due to item response theory methods, but the disadvantage may be that test scores in one content area may not be as comprehensive of a students' academic profile, relative to a composite grade point average across multiple content areas (e.g., English-language arts, science, history/social studies). Moreover, it is important to consider the significance of language in mathematics teaching and learning, specifically for ELLs (Celedón-Pattichis, 2008; Martiniello, 2008; Turner & Celedón-Pattichis, 2011). Considering that mathematics assessments are primarily administered in English in the United States, researchers are responsible for addressing the confoundedness of assessments simultaneously evaluating students' comprehension of mathematical concepts and skills and students' competence in English-language fluency and literacy (American Education Research Association et al., 2014).

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (US). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Capraro, M. M., Capraro, R. M. & Lewis, C. W. (2013). *Improving urban schools: Equity and access in K–12 STEM education.* Information Age Publishing.

Carpenter, T. P., Lindquist, M. M., Matthews, W., & Silver, E. A. (1983). Results of the third NAEP mathematics assessment: Secondary school. *The Mathematics Teacher, 76*(9), 652–659.

Celedón-Pattichis, S. (2008). "What does that mean?": Drawing on Latino and Latina students' language and culture to make mathematical meaning. In M. W. Ellis (Ed.), *Mathematics for every student: Responding to diversity, Grades 6–8* (pp. 59–73). National Council of Teachers of Mathematics.

de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., & Wang, X. (2019). *Status and trends in the education of racial and ethnic groups 2018* (NCES 2019-038). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. https://nces.ed.gov/pubs2019/2019038.pdf

Gándara, P., & Orfield, G. (2010). *A return to the 'Mexican room': The segregation of Arizona's English learners*. The Civil Rights Project/Proyecto Derechos Civiles at UCLA. https://www.civilrightsproject.ucla.edu/research/k-12-education/language-minority-students/a-return-to-the-mexican-room-the-segregation-of-arizonas-english-learners-1/gandara-return-mexican-room-2010.pdf

Gifford B., & Valdés, G. (2006). The linguistic isolation of Hispanic students in California's public schools: The challenge of reintegration. *Yearbook of the National Society for the Study of Education, 105*(2), 125–154. https://doi.org/10.1111/j.1744-7984.2006.00079.x

Gutiérrez, R. (2007). (Re)defining equity: The importance of a critical perspective. In N. S. Nasir & P. Cobb (Eds), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 37–50). Teachers College Press.

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., Rogers, J. E., Jin, Y., & Leinwand, S. (2011). *High School Longitudinal Study of 2009 (HSLS:09): Base-year data file documentation* (NCES 2011-328). U.S. Department of Education, National Center for Education Statistics, National Center for Education Statistics, Elementary/Secondary & Libraries Studies Division. https://nces.ed.gov/surveys/hsls09/pdf/2011328_1.pdf

Ingels, S. J., Pratt, D. J., Rogers, J. E., Siegel, P. H., & Stutts, E. S. (2004). *Education Longitudinal Study of 2002: Base year data file user's manual* (NCES 2004-405). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. https://nces.ed.gov/pubs2004/2004405.pdf

Johnson, R. S. (2002). *Using data to close the achievement gap: How to measure equity in schools*. Corwin Press.

Kao, G., & Tienda, M. (1995). Optimism and achievement: The educational performance of immigrant youth. *Social Science Quarterly, 76*(1), 1–19.

Lippman, L., Burns, S., & McArthur, E. (1996). *Urban schools: The challenge of location and poverty* (NCES 96-184). U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, Data Development and Longitudinal Studies Group. https://nces.ed.gov/pubs/96184all.pdf

Maldonado, S. I., Mosqueda, E., & Bravo. M. A. (under review). Mathematical proficiency practices as predictor of Latinxs' standardized test scores: A simultaneous analysis of student, teacher and school factors.

Martin, D. B., Gholson, M. L., & Leonard, J. (2010). Mathematics as gatekeeper: Power and privilege in the production of knowledge. *Journal of Urban Mathematics Education, 3*(2), 12–24. https://doi.org/10.21423/jume-v3i2a95

Martiniello, M. (2008). Language and the performance of English language learners in math word problems. *Harvard Educational Review, 78*(2), 333–368. https://doi.org/10.17763/haer.78.2.70783570r1111t32

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Cataldi, E. F., Mann, F. B., & Barmer, A. (2019). *The condition of education 2019* (NCES 2019-144). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019144

Milner, H. R. (2012). But what is urban education? *Urban Education*, *47*(3), 556–561. https://doi.org/10.1177%2F0042085912447516

Moses, R. P., Cobb, C. E., Jr. (2001). Radical equations: Math literacy and civil rights. Beacon Press.

Mosqueda, E. (2010). Compounding inequalities: English proficiency and tracking and their relation to mathematics performance among Latina/o secondary school youth. *Journal of Urban Mathematics Education*, *3*(1), pp. 57–81. https://doi.org/10.21423/jume-v3i1a47

Mosqueda, E. (2011). Teacher quality, academic tracking and the mathematics performance of Latino ELs. In K. Téllez, J. N. Moschkovich, & M. Civil (Eds.), *Latinos and mathematics education: Research on learning and teaching in classrooms and communities* (pp. 315–339). Information Age Publishing.

Mosqueda, E., & Maldonado, S. I. (2013). The effects of English language proficiency and curricular pathways: Latina/os' mathematics achievement in secondary schools. *Equity & Excellence in Education 46*(2), 202–219. https://doi.org/10.1080/10665684.2013.780647

Mosqueda, E., Maldonado, S. I., Capraro, R., & Capraro, M. M. (under review). Systematized discrimination: The relationship between students' linguistic minority status, race-ethnicity, opportunities to learn, and college preparatory mathematics.

Murnane, R. J., & Willet, J. B. (2011). *Methods Matter: Improving causal inference in educational and social science research*. Oxford University Press.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological Methodology* (pp. 267–316). American Sociological Association.

Orfield, G., & Lee, C. (2007). *Why segregation matters: Poverty and educational inequality.* The Civil Rights Project at Harvard University. http://www.escholarship.org/uc/item/4xr8z4wb

Perry, T., Steele, C., & Hilliard, A., II (2003). *Young, gifted and Black: Promoting high achievement among African American students.* Beacon Press.

Portes, A., & Rumbaut, R. G. (2001). *Legacies*: *The story of the immigrant second generation*. University of California Press; Russell Sage Foundation.

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

Rosenbaum P. R., & Rubin D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*(387), 516–524.

Rubin, D. B. (1997). Estimating causal effects from large datasets using propensity scores. *Annals of Internal Medicine*, *127*(8, Part 2), 757–763. https://doi.org/10.7326/0003-4819-127-8_part_2-199710151-00064

Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of social composition on academic achievement in high school. *Teachers College Record, 107*, 1999–2045. https://www.tcrecord.org/Content.asp?ContentId=12152

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs* [White paper]. American Educational Research Association. https://www.aera.net/Portals/38/docs/Causal%20Effects.pdf

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Stevens, F. I., & Grymes, J. (1993). *Opportunity to learn (OTL): Issues of equity for poor and minority students* (ED356306). ERIC. https://files.eric.ed.gov/fulltext/ED356306.pdf

Tate, W. (2001). Science education as a civil right: Urban schools and opportunity-to-learn considerations. *Journal for Research in Science Teaching, 38*(9), 1015–1028. https://doi.org/10.1002/tea.1045

Thomas, S. L., & Heck, R.H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education, 42*(5), 517–540. https://doi.org/10.1023/A:1011098109834

Thomas, S. L., Heck, R. H., & Bauer, K.W. (2005). Weighting and adjusting for design effects in secondary data analyses. *New Directions for Institutional Research, 127*, 51–72. https://doi.org/10.1002/ir.155

Trujillo, T. (2016). Restoring the federal commitment to equal educational opportunity: New directions for ESSA's school improvement initiatives. *Education Law & Policy Review, 3*, 141–165. https://educationlaw.org/images/pdf/2016/elpr vol 3.pdf

Turner, E. E., & Celedón-Pattichis, S. (2011). Mathematical problem-solving among Latina/o kindergartners: An analysis of opportunities to learn. *Journal of Latinos and Education, 10*(2), 146–169. https://doi.org/10.1080/15348431.2011.556524

Welsh, R. O., & Swain, W. A. (2020). (Re)Defining urban education: A conceptual review and empirical exploration of the definition of urban education. *Educational Researcher*, *49*(2), 90–100. https://doi.org/10.3102%2F0013189X20902822

Whitehurst, G. J. (2003). *The institute of education sciences: New wine, new bottles* (ED478983). ERIC. https://files.eric.ed.gov/fulltext/ED478983.pdf

---