URBAN INSTITUTE · ELEVATE · THE · DEBATE

Leveraging Nuanced Data to Inform Research and Policy for Immigrant Students and Families

An Essay for the Learning Curve by Kristin Blagg and Marguerite Lukes

June 2022

More than a quarter of US children have at least one immigrant parent, but researchers and policymakers often do not have adequate data on these children's experiences in school, with far-reaching implications for instruction, student support services, and policy.[1] Proxy factors that are reported by school—such as being designated as an English learner or data on race and ethnicity—can tell only part of the story for children of immigrants. This essay demonstrates the untapped value of data on languages spoken at home and explores how policymakers might rethink collection and use of data. Better data on student background can enhance our understanding of students' experiences and provide nuanced information to educators, researchers, and policymakers to better serve distinct student subgroups.

Race remains a salient construct in 21st century US educational contexts, and racism shapes students' lives, yet the seven race and ethnicity categories reported by the US Department of Education do not provide the whole picture.[2] For example, one district in Maine and one in South Dakota have similar percentages of white, Black, and Asian students, but the Maine district enrolls primarily speakers of Somali, French, Arabic, and Chinese, while the South Dakota district enrolls primarily students who speak Spanish, Nepali, German, and Siouan languages.[3] Erasing the nuance of students' experiences can limit policymakers' ability to address students' needs in important ways. In federal civil rights data, for example, Arabic-speaking immigrant Muslim girls from Africa who experience bullying are coded as Black, while their bullying experiences may result from Islamophobic sentiments. Relying on broad panethnic categories risks misunderstanding students' needs and misidentifying solutions.

[1] Cary Lou, Gina Adams, Hamutal Bernstein, and Serena Lei, "Part of Us: A Data-Driven Look at Children of Immigrants," Urban Institute, March 14, 2019, https://www.urban.org/features/part-us-data-driven-look-children-immigrants.

[2] The Elementary and Secondary Education Act (ESEA) requires that states collect achievement and enrollment metrics on all students in publicly funded K–12 schools. The metrics include a set of racial and ethnic subgroups (American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, Black or African American, Hispanic or Latino, white, and two or more races).

[3] "Our Nation's English Learners: What Are Their Characteristics?" US Department of Education, accessed May 31, 2022, https://www2.ed.gov/datastory/el-characteristics/index.html.

500 L'Enfant Plaza SW
Washington, DC 20024
urban.org

All states collect data on the languages that their students speak at home, but few states publicly report these data by school district. Minnesota, however, provides detailed data on the more than 300 languages that the state's students speak, allowing us to use these data to show how subgroups that most policymakers and researchers are familiar with—racial and ethnic groups and English learners—differ from data that indicate immigration background, such as language spoken at home.[4] We also show that including data on languages spoken at home can help explain student academic test outcomes. In sum, data on languages spoken at home provide more nuanced and actionable information that can be used for decisionmaking.

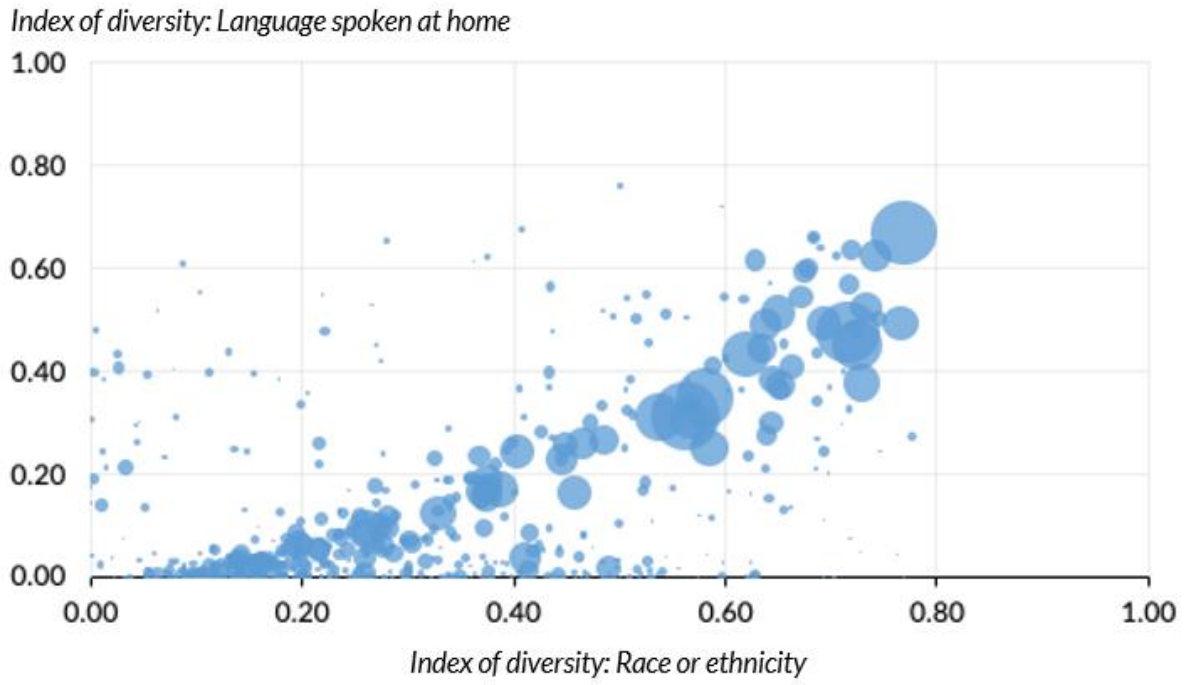
Data on Languages Spoken on Home Reveal Nuance Other Demographic Data Do Not Capture

To examine how the language data compare with district-level information on race and ethnicity, we use Simpson's Diversity Index, where 0 is least diverse (all students belong to one group) and 1 is most diverse. If school districts diverge on their diversity for race and ethnicity versus for languages spoken at home, this may indicate that race and ethnicity cannot fully encompass the diversity of student backgrounds.

[4] Minnesota Department of Education (MDOE), "English Learner Education in Minnesota, 2020-21 Report" (Minneapolis: MDOE, n.d.).

FIGURE 1

Diversity by Race and Ethnicity Only Somewhat Predicts Diversity by Language Spoken at Home



URBAN INSTITUTE

Source: Urban Institute analysis of data from the US Department of Education Common Core of Data and the Minnesota Department of Education.

Figure 1 illustrates the relationship between diversity along these two dimensions. The measures are correlated, but many districts—particularly single-school charter districts—have more linguistic diversity than diversity by race and ethnicity. Similarly, some school districts have a racially diverse population, but students largely speak only English at home.

Data on English Learners Cannot Stand In for Language Data

If nationally reported race and ethnicity data do not fully capture the diversity of immigrant populations, other available data categories, such as the English learner share, better capture student diversity. We explore how well English learner data align with the share of students who come from homes where a language other than English is spoken.[5] If English learner measures are similar and tightly correlated, the extra effort to capture and fully report linguistic data may not be worth it. But we find the opposite; the language data provide different insights than do the English learner data.

[5] The terms "limited English proficient" and "non-English speaking" are still used in data and in state and federal policy but have been largely recognized as being deficit oriented. Terms that have gained wider acceptance among educators and researchers include "multilingual" and "emergent bilingual."

As might be expected, the share of students who are categorized as English learners is correlated with having non-English language speakers at home (appendix figure A.1). But there is still a wide range of English learner shares among districts serving large shares of students from multilingual households, which are primarily charter school districts. And the English learner share is much lower than the non-English language share: about 8 percent English learner statewide versus 16 percent non-English household background. This makes sense because as students learn English, they are declassified from their English learner status. As a consequence, however, their background as a multilingual learner becomes more difficult to access in national datasets.

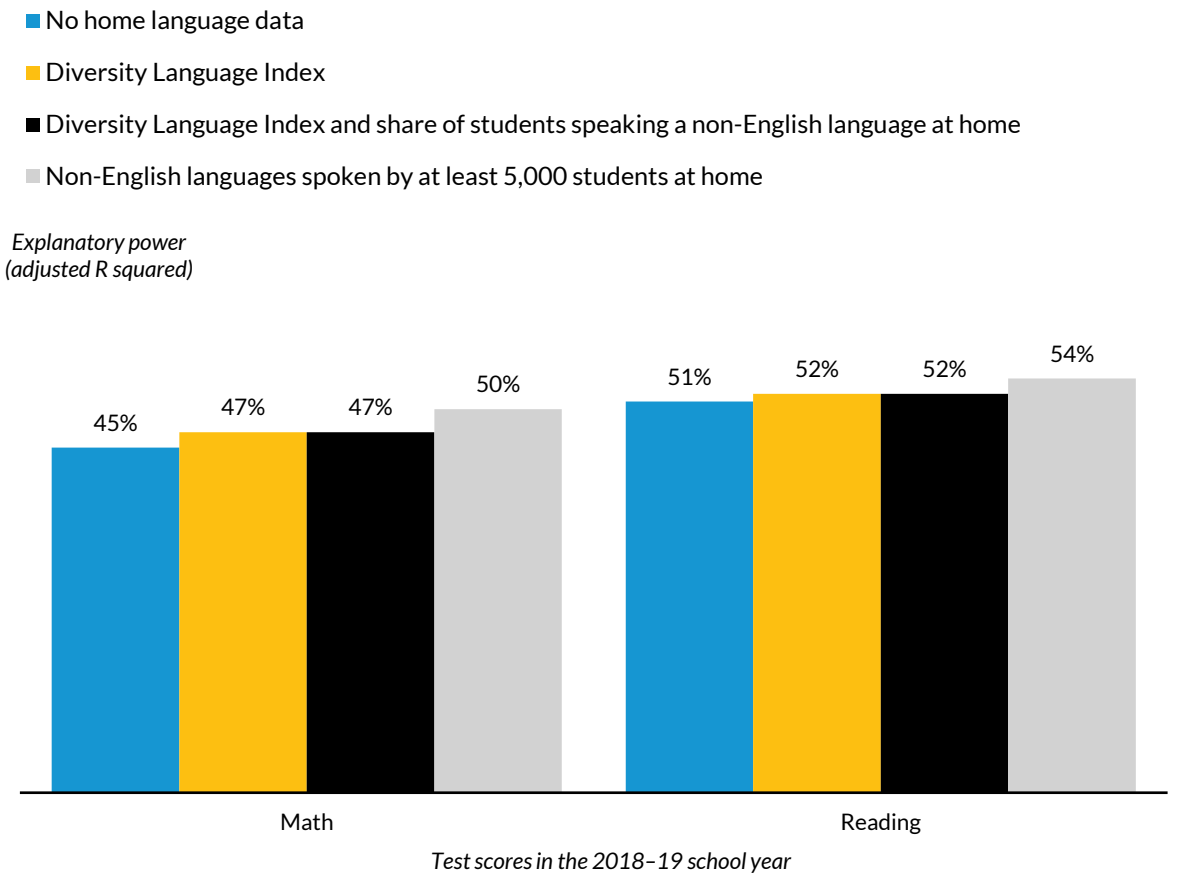
Home Language Data Help Explain Differences in Student Performance

Another way to understand the value of language data might be to see whether they provide additional explanatory power for understanding differences in student test scores. To test this, we looked at the district-level share of Minnesota students who were deemed proficient in math and reading on state tests during the 2018–19 school year (figure 2). We regress the percentage proficiency on each test on a wide array of commonly available demographic indicators, such as the share of students eligible for free and reduced-price lunch, share of students by race and ethnicity, English learner share, and share of students with a disability. Both regressions yield an adjusted R squared value, which is equivalent to the amount of variability in district proficiency levels that appear to be explained by these demographic factors. For math, these variables explain about 45 percent of the variation in proficiency levels across districts, and for reading, they explain about 51 percent.

When we add Simpson's Diversity Index, a variable that describes the diversity of languages spoken in a district, we increase explanatory power slightly, by 2 percentage points for math, and 1 percentage point for reading. For comparison, the magnitude of this effect is similar to, and typically slightly larger than, the explanatory power added when incorporating the share of students with disabilities into a fully specified regression, or adding the English learner share. We find slightly larger increases in explanatory power when adding the share of students who speak a non-English language at home. Finally, we separately incorporate information on languages spoken at home by at least 5,000 students in Minnesota (Arabic, Chinese, English, Hmong, Spanish, Vietnamese, Somali, Oromo, and Karen languages). Accounting for the share of students who speak these languages at home produces a substantial increase in explanatory power.

FIGURE 2

Measures of Language Diversity Increase Explanatory Power on Student Test Score Outcomes



Source: Urban Institute analysis of US Department of Education Common Core of Data, EDFacts, and Minnesota Department of Education data.

Policy Implications

Advocates have long pushed for further disaggregation of ethnic groups to better identify needs that may be overlooked when people from different traditions and cultures are grouped together.[6] Our analysis suggests that publicly reported home language data could inform education policy decisions by contextualizing student achievement. Home language provides a perspective grounded in students' lived experiences and takes communities' unique strengths and challenges into account. All states are required to collect data on home languages spoken by students in the state, and those data follow

[6] See, for example, the work done by the Coalition for Asian American Children and Families: "Invisible No More," Coalition for Asian American Children and Families, accessed May 31, 2022, https://www.cacf.org/policy-advocacy/invisible-no-more.

students even after they have exited English learner status. Public reporting of home languages at the state, district, and school levels could provide much-needed nuance to public data systems.

At the school and district levels, home language data could drive decisions related to translation services, programming for parents, access to multilingual supports for state tests, purchase of instructional materials and tools for student learning, and design of trauma-informed services. At the district and state levels, home language data could inform teacher preparation and hiring, staffing, and achievement reporting. Culturally informed practices could be bolstered by partnering with local organizations that reflect students' backgrounds and recognizing significant holidays in the district calendar (e.g., Lunar New Year or Ramadan).

Districts and schools can use these data to build nuanced understandings of the students they serve, but these data can also provide additional context for state and federal policymakers. Student designations can drive state and national education priorities. For example, some states identify "racially imbalanced" schools,[7] and some districts operate under mandated or voluntary desegregation orders,[8] based on nationally reported race and ethnicity data. To this end, we recommend leveraging home language data to create new designations for schools or districts, such as "linguistically diverse school communities."

We also recommend engaging with diverse stakeholder groups—including members of specific immigrant communities—regarding decisions about data collection and use. States are required to report data on home language to the federal government each year, yet those data are rarely made public.[9] Making data public at the district level would protect student privacy while adding much-needed detail on student subgroups who may be otherwise hidden in the data. Providing national public reporting of language categories will also help researchers and policymakers more fully understand school contexts and would help advocates and educators by providing nationally comparable data on the students they serve.

The intersection of different student demographic identities is complex. Our work highlights the potential value of these data but raises important questions about how these data can be responsibly shared and used. All states gather these data, but few report them. Only Arkansas, California, and Minnesota publicly report languages spoken at home at the district level, and just five other states appear to report any home language data, even at the most aggregate level, in their state reporting. An exploration of how and why states gather and report these data and how the data have informed local and state policy could bring us closer to shedding some light on issues that have remained largely

[7] States that use this school-level identification include Connecticut (see "Racial Imbalance: Overview," Ct.gov, accessed May 31, 2022, https://portal.ct.gov/SDE/Legal/Racial-Imbalance) and Massachusetts (see "Racial Imbalance Advisory Council," Massachusetts Department of Elementary and Secondary Education, accessed May 31, 2022, https://www.doe.mass.edu/bese/councils/racial.html).

[8] Yue Qiu and Nikole Hannah-Jones, "A National Survey of School Desegregation Orders," ProPublica, December 23, 2014, https://projects.propublica.org/graphics/desegregation-orders.

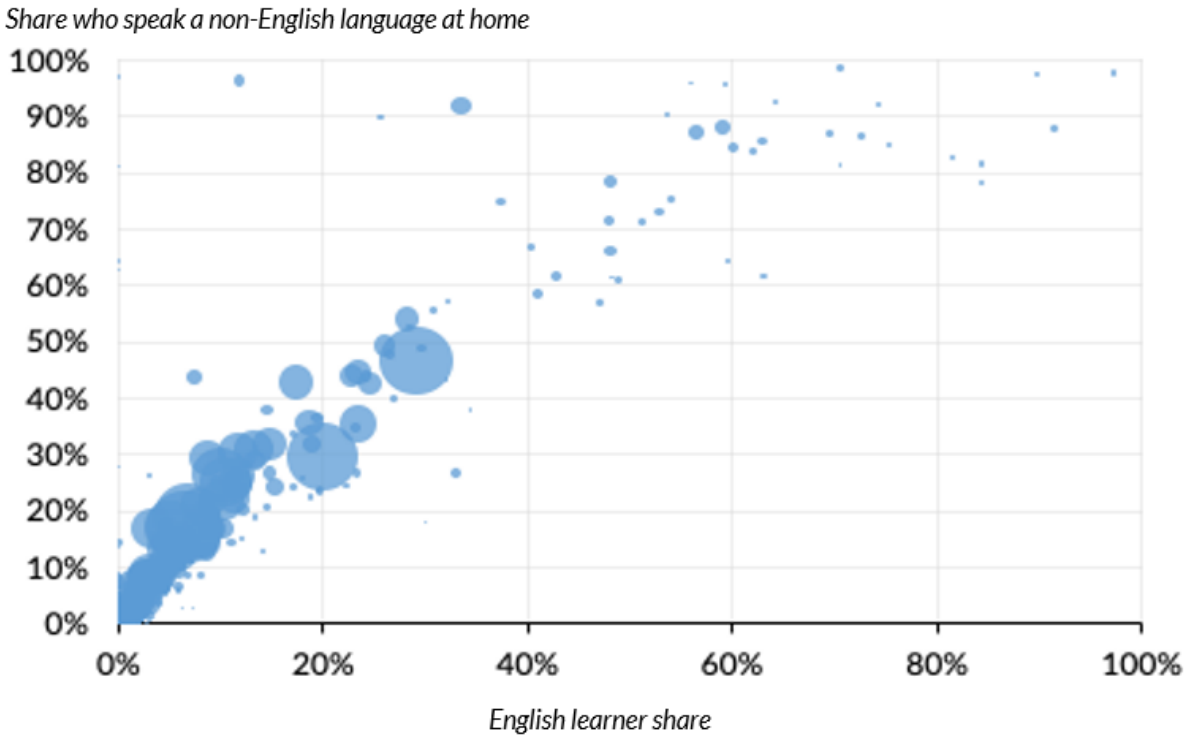
[9] States use the *Consolidated State Performance Report* (CSPR), a federally mandated annual reporting tool, to report their metrics directly to the US Department of Education. In addition, states are required to report publicly some of these metrics in their Annual State Report Card required under Section 1111(h)(1)(A) of the ESEA.

overlooked and unexplored. Further research in this area would necessitate the inclusion of stakeholders from linguistically and ethnically diverse communities most affected by these policies.

Appendix

FIGURE A.1

English Learner Share Is Closely but Not Perfectly Associated with Diversity by Language Spoken at Home



Source: Urban Institute analysis of data from the US Department of Education Common Core of Data and the Minnesota Department of Education.

Kristin Blagg is a senior research associate in the Center on Education Data and Policy at the Urban Institute. Marguerite Lukes is director of research and innovation at Internationals Network.

Acknowledgments

This essay was funded by the Walton Family Foundation and the Bill & Melinda Gates Foundation as part of the Learning Curve essay series. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the author and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at www.urban.org/fundingprinciples.

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org