Non-English Language Students in Palm Beach School District: An Examination of Administrative and American Community Survey Data

Robert Kominski , Tiffany Julian
Social, Economic and Housing Statistics Division
U.S. Census Bureau

ABSTRACT

The American Community Survey (ACS) has been envisioned by the Census Bureau as the replacement of the once-a-decade long form data collection previously undertaken as part of the decennial Census. In addition to providing more frequent and timely data, the survey is designed to provide reliable data at small geographic levels for meeting policy and programmatic needs. This analysis examines data from the Palm Beach Florida (PBF) school district regarding school children who speak a language other than English. We compare estimates from the administrative school-based data with those from the sample survey estimates of the ACS. We find that the ACS one-year datafile provides reasonable quality estimates for the major language groups. Five-year ACS data provides better information for more specific languages, as well as better specific estimates that are closer to the administrative levels.

Introduction

One of the greatest needs researchers and policy-makers have is for timely data about detailed characteristics of the population. As social, demographic and economic issues emerge and change, the need for high quality reliable data on an ongoing basis is a central need, both for the research community as well as for the legions of practitioners who must often take prompt action for changes in their local environments.   For many decades, these researchers have used administrative data, local-scale surveys, and often, summary detailed data produced by the Census Bureau as part of the once-a-decade long form Decennial Census data collection.

The main criticism of the long form data was that within a few years of collection (in the 'zero year'), the small-scale geographic estimates could already be out-of-date and misleading. While estimates of the base population count are annually updates by the Census Bureau as part of its Population Estimates Program (including fundamental demographic characteristics such as sex, race and age), estimates of a myriad of other social, demographic , economic and housing phenomenon go unmeasured.

This very basic limitation of the long form data became readily apparent in the 1970's and 1980's, as growing demand for more and more extended analyses of data came to the fore. Driven by this demand/need of the data community, as well as other fundamental logistical and operational challenges associated with a massively growing decennial census operation, the Census Bureau made the decision to decouple the long form data collection from the Decennial Census and implement it as an independent survey operation.   This new program, which has come to be known as the American Community Survey (ACS), began test operations in the late 1990's, was functional enough to provide basic estimates for comparison/comparability concurrent with 2000 Census, and was fully implemented (including groups quarters and Puerto Rico) in 2006.

Fuller explanation and description about the design and functional components of the program are discussed in a wide variety of sources (CIT).   But in short, the survey now delivers estimates comparable

to those from the decennial census long form, but does this every single year. And, utilizing the multi-year period files (which constitute data collections from 5 sequential years of data), also provides hundreds of thousands of unique substantive estimates down to the Census tract and block group geographic level.

One of the concerns expressed about the ACS is the ability to provide small-scale geographic estimates of comparable statistical quality to those from the traditional long form-based estimates. While both the Census Bureau and other researchers have provided the start of a sizable base of knowledge on this topic (CIT), the final judgment on this topic is far from being issued or known, and there is no clear consensus that the ACS small-scale geographic estimates are sizably deficient from their predecessors in the census long form. In fact, from a perspective of timeliness and periodicity, the ACS estimates are VASTLY superior.   From a substantive standpoint, the ACS provide far more data estimates and capability for analyzing emerging and evolving issues than any other national general survey.   Within the Federal statistical system, it has rapidly moved to become the flagship survey data collection program for a wide variety of data and information needs.

This analysis continues the process of providing real empirical evidence to demonstrate support for the statistical and substantive validity of the ACS program. Estimates from the Palm Beach Florida School District are compared to those produced from the ACS in an effort to assess the quality and properties of the two datasets. In this paper we examine the estimates, their statistical reliability, and the degree to which they substantively corroborate or conflict one another.   The specific estimates in question are the counts of school-age children who speak a language other than English, and the specific non-English language they speak.

Data and Analysis Questions

This analysis uses data from both the one-year and five-year data files from the ACS as well as administrative data provided to the Census Bureau by the Palm Beach Florida (PBF) school district. School enrollment and language use items are collected as part of the ACS, and are reproduced in Appendix A. The language question is a 3-part question in use since the 1980 census. The question asks of all persons ages 5 above: (a) if they speak a language other than English at home; (b) what that language is; (c) "how well" they speak English.   In many households, answers to all or most questions for all persons may be provided by one 'household respondent'. The 2010 one-year dataset is designed to provide estimates of areas of populations 65,000 or more, and the PBF school district, at 178,301 students, easily exceeds that threshold. These data have the advantage of being timely data—collected in the same year as our administrative records.

The five-year dataset has the advantage of having a much larger sample, and therefore has the possibility of identifying subpopulations that may be relatively rare.   The five-year data are the result of combining data collected over five consecutive years so that the sample size is much larger than any one year of data. While the single-year data for PBF (county) is based on 7800 households, the five-year sample constitutes 40,281 households.   While the five-year estimates benefit from the large sample size, they lack timeliness. When a population changes between years, the resulting estimate can be seen as the average over the whole time period. For small but stable populations that don't change much over time, the estimates should be much better with the five-year data

The administrative data were provided to Census by the PBF school district and reflect their counts of public school attendees who report having non-English language backgrounds. The specific questions asked are: (1) Is a language other than English used in the home; (2) Does the student have a first language other than English; (3) Does the student most frequently speak a language other than English. Note than none of these questions match exactly those asked in the ACS, so direct comparability is an issue in the analysis.

This analysis examines several issues regarding non-English language students in the PBF school district. We hypothesize that the one-year ACS data will provide reasonable estimates of the overall level of non-English language students in the PBF school district. We also hypothesize that the five-year ACS dataset will provide more complete and more reliable estimates of a greater number of specific languages spoken as well as a better sense of the relative size of these language groups.

Results

Table 1 provides estimates of the count and relative percentage distributions of the languages spoken by children in the PBF school district, based on the administrative counts for school year 2010, the one-year (2010) ACS dataset and the five-year (2006-10) ACS dataset. The table has been sorted by the number of speakers recorded by the school district.

The summary results of this detailed table are shown below:

	PBF admin	2010 ACS	2006-10 ACS
Total public K-12	178,301	176,205	171,268
English only	125,679 (70.5%)	111,619 (63.3)	116,316 (67.9)
Non-Eng speakers	52,622 (29.5)	64,586 (36.7)	54,952 (32.1)
# of languages	144	33	57

According to student registration forms, the Palm Beach County Schools enrolled 178,301 students in Kindergarten through 12th grade in 2010, and 29.5% of those students reported speaking a language other than English. There were 144 different languages reported by students to the school district- 39 of these languages were reported by just one student each.

The ACS 2010 single-year estimates do a good job of estimating public school enrollment--176,205 students (+/-6,240). The actual enrollment reported by the school district falls nicely within the ACS estimate and is off by about 2,000 students.
The single-year estimates are based on a sample of 7800 households representing over 500,000 households in the county. With 144 different languages spoken in the school district, the data are limited in their ability to capture the diverse number of languages spoken by students. The 2010 ACS estimates that 36.7% of public school children spoke a language other than English at home and these children spoke 33 different languages. The ACS estimate for non-English languages is about 7% (12,000 more students) higher than the school district number, while the number of reported languages is much

lower (111 less languages).   As mentioned earlier, ACS single-year estimates are best for large populations (at least 65,000 population) but we are now examining a much smaller sub-population— children enrolled in public school grades K-12 who speak a language other than English with a known population of about 56,000.

Using the 5-year ACS data file for 2006-2010, the estimate for public school enrollment in Palm Beach County Schools is 171,268 students (+/-2,535). This is off by about 7,000 students and the difference does not fall within the margin of error. Looking at the percent who speak a language other than English, we see better news. ACS 2006-2010 estimates that 32% of enrolled students spoke a language other than English—much closer to the 30% reported by the school district (2,300 more students). This dataset also captured 57 distinct non-English languages , which is 24 more than the single-year data did.

Despite the large number of distinct languages captured, the same four: Spanish, French Creole, Portugese and Vietnamese, account for respectively, 89.6 (admin), 85.2 (ACS1yr) or 87.5(ACS 5yr) percent of all non-English language students in the Palm Beach school district.

Specific Languages Spoken

The PBF administrative records indicate that 18% of students spoke Spanish. ACS 2010 estimates that 21% spoke Spanish compared to 20% in 2006-2010. Put another way, the single year data over-estimates 6,193 Spanish speakers while the five-year data over-estimates by only 2,263. The next largest language reported by the school district is Haitian Creole/French Creole (8%). In 2010 the estimate is 9% and in 2006-2010 the estimate is 7%. It is notable that the ACS estimate of standard French, Patois, and Jamaican Creole are much higher than the school district reports. Because these languages are all related, we can look at a combined estimate of these languages. The school district reported 14,416 speakers of these languages compared to 18,710 in the single-year file and 13,630 in the five-year file. The reported number and the 5-year estimate differ by 786 students or .1%. The next largest language reported by the school district is Portuguese (1%). Both ACS files also estimate 1% of students speak Portuguese. In this case the single-year file underestimates by 556 (.3%) students while the five-year file overestimates by 329 (.2%) students.

Overall, ACS does a very good job at estimating larger language populations. Single-year estimates seem to provide a timely and accurate estimate of school enrollment in the school district while the five-year estimates seem to provide better estimates of detailed language use. In a county with a population of over 1.3 million people, ACS is able to estimate the number of public school children who speak non-English languages to within a few hundred.

The sort of Table 1 also gives us a better idea of the level of consistency between the school district counts and those from the ACS.   In general, the 5-year ACS data seem to have better consistency with the admin estimates than the one-year ACS data. Several languages which show large relative +deviations in the 1-year ACS data come much closer into line using the 5-year files, for example, Gujarathi (A:107, ACS1: 1189, ACS5: 235), Turkish (68; 457;185) or Cantonese (34; 579; 172).

Another type of inconsistency occurs when the admin data reveal few or no speakers, yet the ACS shows speakers in abundance. Examples of this include: Burmese (11; 161; 59), Ukranian (9, 176, 11) Bengali (1; 440; 233) and Mande (0; 128, 48).

Further work will focus on whether this may be a function of misclassification of the languages on either the admin or ACS side.

English Speaking Ability

We were also able to obtain data from Palm Beach County Schools on English Language Learners. Unfortunately, these data included persons in Adult Learning Centers. We eliminated all schools with "ADULT" in the name but the data probably stil includes some persons who are not traditional K-12 enrollees.   For example, there were 1625 English Language Learners in the dataset who are reported to speak English Only. There are also 6 reported languages that were not listed in the original file (Gaelic, Kazakh, Tatar, Malay, Berber, Setswana) and 2 languages with more English Language Learners than speakers (Uzbek, Azerbaijani).   Our analysis will expand to examine these data as well. At this point, one of our main concerns is that there may not be sufficient comparability between the administrative data and the ACS survey data to evaluate whether the estimates are comparable or not.