POPULATION DIVISION

Evaluation of the 1990 School District Level Population Estimates Based on the Synthetic Ratio Approach

by
Esther R. Miller

Working Paper Series: No. 54

September 2001

WORKING PAPER NO. 54

Evaluation of the 1990 School District Level Population Estimates Based on the Synthetic Ratio Approach

by
Esther R. Miller

Population Division
U.S. Census Bureau
Washington, DC 20233-8800

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

ABSTRACT

The Census Bureau was tasked with conducting research and evaluation and developing a methodology to produce updated estimates of the total population and the total number of school-age children in each school district. This paper provides an overview of the methodology and limitations, the steps necessary to create the synthetic population estimates, problems we encountered, and results from our evaluation of the data.

Acknowledgments:

The author would like to thank Paul Siegel (Small Area Income and Poverty Estimates), Bashir Ahmed and Signe Wetrogan (Population Division) for their invaluable comments and contributions to this research.

Evaluation of the 1990 School District Level Population Estimates Based on the Synthetic Ratio Approach

I. INTRODUCTION

The elementary and secondary schools in the United States depend on federal dollars to supplement programs for disadvantaged children. Title 1 of the Elementary and Secondary Education Act provides a means for the Department of Education (DOE) to distribute federal funds to school districts.

Prior to School Year (SY) 1997/1998, the distribution of federal dollars to school districts was carried out in a two-step process. First, the DOE allocated federal dollars to counties. States then had the responsibility to distribute the federal dollars to school districts. In order to determine the amount of money to allocate to a state, the DOE used the most recent decennial data on the number of school-age children in poverty in each county within the state. States then used a variety of data sources to allocate the monies down to the school districts including special decennial census tabulations of the number of school-age children in poverty in each school district.

In 1994, Congress enacted a law authorizing the Department of Education to allocate Title 1 funds directly to school districts, beginning with school year 1997/1998. In doing so, Congress also specified that the DOE use updated estimates of the number of school-age children in poverty in each school district rather than the once-a-decade measures from the decennial census.

The Census Bureau was tasked with conducting research and evaluation and developing a methodology to produce updated estimates of the number of school-age children in poverty. Because the distribution of the funds also requires updated estimates of the total population and the total number of school-age children in each school district, the Census Bureau also had to develop methodologies for these data requirements.

This paper focuses on the development and evaluation of the methodologies to produce updated estimates of the *total* population and the total number of *school-age children* in each school district. It is divided into five sections. Section I is the introduction. Section II describes the methodology developed to produce the updated population estimates for school districts and issues that affect the production and subsequent accuracy of the estimates; Section III describes the methodology used to evaluate the school-district estimates; Section IV presents the results of the evaluation; and Section V presents conclusions and discusses plans to improve the population estimates for school districts. A discussion of the development and evaluation of the

methodology to produce updated estimates of the number of school-age children in poverty is presented in a separate paper.[1]

II. DEVELOPMENT OF METHODOLOGY

This section presents an overview of the methodology used to produce the estimates of the total population and the school-age population in each school district.

As noted in the prior section, the Census Bureau was tasked with developing the methodology to produce updated estimates of the total population and the school-age population in each school district. To comply with the legislation, the methodology had to be developed and implemented for the allocation of funds for school year 1997/1998.

Although the Census Bureau did have a program to develop and produce annual estimates of the population of functioning governmental units, the methodologies developed for those estimates could not be used to produce updated estimates of school districts. Therefore, it was necessary for the Census Bureau to construct a new methodology to produce the population estimates for school districts.

Factors Affecting Development of Methodology

In developing the methodology, we encountered a number of factors which complicate the development of estimates for school districts.

School Districts are Small with Unique Boundaries

School districts are small with unique boundaries. As such, little Census or other data are available as input to an estimation methodology. In 1990, there were 15,226 school districts in the United States. Table 1 shows that approximately 50 percent of these school districts have a total population of less than 5,000 people. Approximately 82 percent of all school districts have an estimated total population of less than 20,000 people (U.S. Census Bureau, 1997).

[Table 1 about here]

In most parts of the United States, school district boundaries are unique in that they do not coincide with other governmental units for which data are regularly tabulated. There are only seven states where school district boundaries coincide with county boundaries, accounting for only 928 of the 15,226 school districts in the United States. Although most school districts are confined to a single county, some cross county boundaries, further adding to the complications in developing an estimation methodology.

School Districts are Defined by Relevant Grades

[1]See http://www.census.gov/hhes/www/saipe/schooltoc.html for the documentation.

School districts are defined according to the grade levels served by the school district. Therefore the estimates of the number of school-age children in each school district had to be calculated according to the grade level served by the school district. In 1990, about 74 percent of the school districts across the United States served grade levels kindergarten through 12th grade. The remainder of the school districts served only specific grades such as kindergarten through 6th grade (22 percent) or 9th through 12th grade (4 percent).[2]

For those school districts which served only partial grade levels, it was necessary to translate the grade levels served back to relevant ages. The 1990 census data on highest grade completed together with data from the October supplements of the 1988, 1989, and 1990 Current Population Surveys provided the necessary information to develop a grade to age relationship.

The translation of grade to age was done so that each school-age child could be assigned to one and only one school district. Thus, the sum of school-age children across school districts would equal the total number of school-age children in the United States. However, this is not true for the sum of the total population across school districts. Because a school district may provide elementary grade service on the same piece of land as a district that provides education for middle school grades, the estimates of the total population for these overlapping school districts will be double counted. Thus, the sum of the estimates of the total population for all school districts cannot be compared with the total population of the United States.

School District Boundaries Change Over Time

Several changes may occur to school districts over time. School districts can annex new territory over time; school districts can close; and new school districts can be created. In order to maintain correct and up to date boundaries, the Census Bureau must periodically survey school districts to obtain current boundary information. Additionally, the changes to boundaries complicate the complete evaluation of any methodology.

Choosing the Ratio Methodology

The complexities outlined above and the scarcity of data available for school districts led the Census Bureau to choose a ratio or synthetic approach to produce the school district estimates. In choosing the ratio approach, the Census Bureau decided to rely upon the 1990 census to provide a starting point and the annual estimates of the county population to provide the basis for change. The annual estimates of the total population for counties would provide the basis for change in the *total population* for school districts. The annual estimates of the population by age for counties would provide the basis for change in the *school-age population* for school districts. This approach assumes that all school districts within a county change at the county rate. The formula for developing the estimates for the post 1990 period is:

$$P_{(sd\ t)} = P_{(sd\ 1990)} / P_{(county\ 1990)} * P_{(county\ t)}$$

where:

[2]Special tabulation by the U.S. Census Bureau.

$P_{(sd\ t)}$ = Estimated school district population in current boundaries for time t
$P_{(sd\ 1990)}$ = School district population in current boundaries from 1990 census
$P_{(county\ 1990)}$ = County population from 1990 census
$P_{(county\ t)}$ = County population for time t

While most school districts are confined to a single county, some do cross county boundaries. For those cases where the school district crosses county boundaries, it is necessary to construct a separate ratio and separate estimates for the school district piece in each county. In these cases, as a final step, the separate school district county pieces are summed to produce the school district estimate.

Assumptions Associated with Ratio Approach

The ratio approach assumes that the ratio of the school district population to the county population will remain constant over time. In other words, it assumes that the population in each school district county piece changes at the same rate as that of the county. However, in reality this may not be the case. If the county population is estimated to decline, but the school district population in that county increases or vice versa, the resulting estimates of the school district population will be biased.

The estimate is further complicated when a school district crosses county boundaries. In that case, the ratio method assumes that each school district-county piece grows at the rate of that county. In a school district that crosses county boundaries, one of the counties it comprises may see a population spurt whereas the other county may experience a decline in population. When the two county pieces are summed together, the school district population may be underestimated or overestimated, depending upon the size of the school district pieces.

III. EVALUATING THE RATIO APPROACH

Development of Ratio Estimates for Evaluation

To do a complete evaluation of the school district methodology, we need to have school district data at two points in time. The data for the 1980 and 1990 censuses provide us with that opportunity. To evaluate the ratio methodology, we used the 1980 census as the base, developed an estimate for 1990, and compared the estimate to the 1990 census data. The estimates were produced for both the total population and the school-age population aged 5-17 years. For this evaluation, we developed four sets of synthetic population estimates.

Set 1: County Estimates-Based Model

To evaluate the ratio approach applied to an estimate of the county population (as would be the case in the post 1990 period), we must develop a 1990 estimate for the county. For this test, we

used the 1990 estimate of the county population that had been developed using our standard county estimates approaches and based on the 1980 census.[3]

To produce these estimates, we first compute the ratio of the school district population to county population using the 1980 census data. Then we apply the ratio to the 1980-based estimate of the 1990 county population developed by the Census Bureau. This evaluation measures the effect of the ratio approach as well as any error caused by the estimate of the county population.

$$P_{(sd\ 1990)} = P_{(sd\ 1980)}/P_{(county\ 1980)} * P_{(county\ 1990)}$$

where:

$P_{(sd\ 1990)}$ = Estimated school district population in 1990
$P_{(sd\ 1980)}$ = School district population from 1980 census
$P_{(county\ 1980)}$ = County population from 1980 census
$P_{(county\ 1990)}$ = Estimated county population in 1990

Set 2: County Count-Based Model

This approach is very similar to Set 1 except that the ratios are multiplied by the 1990 census data for the county population rather than the 1980-based estimate. We are assuming that all school districts within the county change at the same rate as the county. Although for the post 1990 period we would only have estimates data available, this estimate is a good benchmark against which to judge all other model-based estimates.

In this approach, we multiply the ratio of the 1980 school district population to 1980 county population by the 1990 census county population.

$$P_{(sd\ 1990)} = P_{(sd\ 1980)}/P_{(county\ 1980)} * P_{(county\ 1990)}$$

where:

$P_{(sd\ 1990)}$ = Estimated school district population in 1990
$P_{(sd\ 1980)}$ = School district population from 1980 census
$P_{(county\ 1980)}$ = County population from 1980 census
$P_{(county\ 1990)}$ = County population from 1990 census

Set 3: State Growth-Based Estimates

This approach is similar to Set 2 except that it assumes that the school districts all change at the same rate as that of the state. To develop the estimates, we multiply the ratio of the 1990 state population to 1980 state population by the 1980 school district population.

$$P_{(sd\ 1990)} = P_{(State\ 1990)}/P_{(State\ 1980)} * P_{(sd\ 1980)}$$

[3] See http://www.census.gov/population/methods/stco99.txt for the methodology.

where:

$P_{(sd\ 1990)}$ = Estimated school district population in 1990
$P_{(State\ 1990)}$ = State population from 1990 census
$P_{(State\ 1980)}$ = State population from 1980 census
$P_{(sd\ 1980)}$ = School district population from 1980 census

Set 4: National Growth-Based Estimates

This approach is also similar to Sets 2 and 3 except that it assumes that the school districts all change at the same rate as that of the entire United States. To develop this estimate, we multiply the ratio of the 1990 national population to 1980 national population by the 1980 school district population.

$$P_{(sd\ 1990)} = P_{(National\ 1990)}/P_{(National\ 1980)} * P_{(sd\ 1980)}$$

where:

$P_{(sd\ 1990)}$ = Estimated School district population in 1990
$P_{(National\ 1990)}$ = National population from 1990 census
$P_{(National\ 1980)}$ = National population from 1980 census
$P_{(sd\ 1980)}$ = School district population from 1980 census

Note that the assumptions underlying the models may not be realistic. For example, the population growth in a school district does not correspond to the growth in a county or state. Similarly, it is not reasonable to assume that each and every school district will grow at the same rate as the nation.

Creating a Comparable Universe of School Districts Across the Decade

To do a complete evaluation of the methodology, we need a comparable universe of school districts over the 1980 to 1990 time period. Optimally, for our analysis we would use a matched 1980 and 1990 file, geocoded to identical school district boundaries. The advantage of this type of file is that we would not need to make assumptions about school district boundaries across the decade.

If the Census Bureau had a 1980 data file geocoded to the 1990 school district geography we could simply apply synthetic ratios to 1990 census data and compare the expected value to the "truth" in 1990. If we were able to geocode 1990 data into 1980 school district geography, we could administer the same approach. However, neither data set is available.

Considering we do not have files geocoded to the same boundaries, we concluded we needed to prepare a universe of school districts that are "equivalent" across the decade. The starting point for our universe is the total number of school districts in 1990 (15,226). (See Table 2). We first excluded 928 school districts that were coterminous with county boundaries as the stable shares approach perfectly predicts the population for the 1990 school district for this set of school districts.

[Table 2 about here]

Essentially, we could apply the synthetic ratio approach to the remaining 14,298 districts. However, in order to have an "equivalent" universe file over the decade, we also removed:

1. School districts with limited grade ranges[4] (4,018);
2. School districts which were newly formed between 1980 and 1990 (416);
3. School districts in counties where the county boundaries changed between 1980 and 1990 (12)[5];
4. School district county pieces did not match up across the decade (609); and
5. School districts with a population size of less than 31 people (42).

The final universe for the 1980-1990 evaluation file contained 9,201 matched school district identification numbers.

Evaluation Measures

To compare and evaluate the estimates, we used two standard statistical measures: (1) the Mean Absolute Percent Error (MAPE), and (2) the Mean Algebraic Percent Error (MALPE).[6] The MAPE is computed as the sum for all school district pieces of the absolute difference between the estimate and the 1990 census figure divided by the number of school districts. The MAPE measures the accuracy of the estimates. The MALPE is computed in a similar manner, except that we take the sign of the difference into consideration. Positive mean algebraic percent errors indicate overestimation of a population and negative errors indicate an underestimate of a population.

We also examined weighted MAPEs. The unweighted statistics treat each school district with equal importance, regardless of size. The weighted MAPEs, on the other hand, take into

---

[4] Most school districts cover the grade range of K-12. These are known as unified school districts. A non-unified school district does not cover grades K-12 but instead covers elementary, middle, or high school grades. If a school district is not unified across the decade, it is not possible to determine whether the grades the district includes are the same across time (U.S. Department of Education, 1999).

[5] We assumed the school district boundaries did not change if the identification number did not change over the decade. This assumption may not always be correct because the state did not always assign new IDs when land was annexed over the decade, political boundaries changed, etc. (U.S. Department of Education, 1999).

[6] See Appendix for the formulas for school district estimators and evaluation statistics for the models. The appendix includes references to both population and poverty estimates. There are some slight differences in the terminology. Our text refers to MALPE whereas the appendix refers to MALP. Additionally, Model-based refers to our Set 1, census county-based refers to our Set 2, and the naive-based refers to our growth-based estimates. Thanks to William R. Bell for providing the statistical explanation for the computations (U.S. Census Bureau, 1998).

consideration the size of a school district, measured by the total population or the school-age population in that school district. Weighting by the total population in each school district addresses the size of the school district population affected. Weighting by the number of school-age children indicates how accurate the estimates are for the districts containing the average child.

IV. RESULTS OF THE EVALUATION

For purposes of this evaluation, we developed four sets of synthetic population estimates. Set 1 uses the ratio approach and the 1980 based county population estimate. Set 2 is similar except that it uses the 1990 census data for the county rather than the 1980 based estimate. The differences between Set 1 and Set 2 represent the additional error in the ratio approach introduced by using an estimate of the population rather than the census counts. Sets 3 and 4 represent alternatives to a county-based approach. Set 3 assumes that the school district grows at the same rate as that of the state, while Set 4 assumes that the school districts all grow at the national rate.

Overall Quality

As shown in Table 3 and Figure 1, the county count-based estimates have the smallest unweighted MAPEs (12.6 and 16.0), followed by the county estimates-based (13.3 and 16.9), the state growth-based (16.4 and 18.9), and national growth-based estimates (18.9 and 20.6). This pattern holds both for total population and school-age population aged 5-17, whether the MAPEs are weighted or unweighted.

[Table 3 and Figure 1 about here]

Table 4 presents the results of comparing the MAPEs across each set of estimates. As show in the first row of Table 4, we lose only a minor amount of accuracy when we use an estimate rather than the census count as the base for the 1990 county data. Comparing Set 1 to Set 3 and Set 4 indicate that the use of the ratio approach at the county is superior to one that uses state or national growth rate assumptions.

[Table 4 about here]

Using the MAPEs as our unit of analysis, we would conclude that the Set 2 approach is the most accurate for estimating the school district population. However, the Set 2 (county count-based approach) can be produced only at the census year. Therefore, if we must rely on the synthetic approach, we need to employ a set of estimates. And as shown by the comparison to Sets 3 and 4, the use of the county estimate is superior to a method that uses state or national growth rate assumptions. For this reason, the remainder of this section reports results from the county estimates-based MAPEs and MALPEs.

Quality of the Estimates by Demographic and Economic Characteristics

To evaluate the amount of "bias" or other patterns in the county estimates-based school district estimates, we selected ten economic and demographic characteristics. These characteristics are a subset of those the National Academy of Sciences used to evaluate poverty estimates at the county level.[7] The ten characteristics are:

1. Size of the School District in 1980;
2. Size of the School District in 1990;
3. Population Growth, 1980-1990;
4. Percent Poor School-age Children in 1980;
5. Percent Poor School-age Children in 1990;
6. Numerical Change in Poverty Rate for Children, 1980-1990;
7. Census Division;
8. Percent Hispanic in 1980;
9. Percent Black in 1980; and
10. Percent Group Quarters in 1980.

Table 5 shows both the unweighted and weighted MAPEs and unweighted MALPEs[8] for total population, by the selected characteristics. Similarly, Table 6 shows the unweighted and weighted MAPEs and unweighted MALPEs by characteristics for school-age population aged 5-17. Additionally, the two tables present the total population (or school-age population) and the percent of the population in each category.[9]

[Table 5 and Table 6 about here]

---

[7] See National Research Council, 1998.

[8] We will not discuss weighted MALPEs because the sum of the MALPEs for each economic or demographic characteristic would be equivalent to zero if all of the school districts in each county were represented in our sample, thus the weighted MALPEs are meaningless to the analysis.

[9]The unweighted number of school districts in each category of the demographic and economic characteristics remain the same across Table 5 and Table 6. This is because the demographic and economic categories (e.g., Size of the School District in 1980 or Percent Poor School-age Children in 1980) were defined based on the characteristics of the total population in a school district. For example, if the total population in a school district is 9,000 and the school-age population in a school district is 4,500 the school district falls into the school district population of 5,000 - 9,999. In Table 5, the total population is determined by weighting the number of school districts by the total population in each school district. In Table 6, we determined the school-age population by weighting the number of school districts by the number of school-age children in each school district.

Figures 2 through 11 are pictorial representations of the weighted and unweighted MAPEs for both the total and school-age population, by demographic and economic characteristics.

Size of the School District in 1980 (See Figure 2 and Tables 5 and 6)

- The unweighted MAPEs for school districts with fewer than 5,000 people are almost two times as high as the MAPEs for all other population categories (the MAPE for total population is 18.1 and the MAPE for the school-age population is 22.8). Whereas the weighted MAPEs for the same set of school districts are one and half time as high as the other categories.

- For small school districts (those with a population less than 5,000), we overestimated the total population and the school-age population by 8.8 percent and 6.9 percent, respectively (see MALPEs).

- For larger school districts (those with a population more than 40,000), we overestimated the total population and the school-age population by 3.2 percent and 5.0 percent, respectively.

- Almost half of all districts (48.2 percent) have a total population of less than 5,000 people. However, these districts account for only 6.9 percent of the total population and 7.7 percent of the school-age population.[10]

- School districts that are populated by 20,000 or more people represent 16.5 percent of all school districts, but are populated by two-thirds (66.6 percent) of the total population and about two-thirds (64.5 percent) of the school-age children.

[Figure 2 about here]

Size of the School District in 1990 (See Figure 3 and Tables 5 and 6)

- The relationship between the size of the district and the size of the MAPEs in 1990 show the same patterns as what we see in 1980.

[10] The findings in the last two bullets above are consistent with findings shown in Table 2 in that about one half of all school districts are made up of less than 5,000 people. The difference is that Table 2 is based on the total number of school districts as of 1989-1990 (15,226 school districts); whereas the evaluation universe is based on 9,201 districts.

- With the exception of the smallest school districts, the MALPEs based on population size in 1990, are lower than those based on the school district population size in 1980.

[Figure 3 about here]

Population Growth, 1980-1990 (See Figure 4 and Tables 5 and 6)

- When the size of the school district increases or decreases over the decade by 10 percent or more, the MAPEs are much higher as compared with school districts with a stable amount of growth.

- For school districts that lost 10 percent or more of their total population during the 1980 to 1990 period, the ratio approach tended to overestimate their total population by an average of 28.9 percent.

- Conversely, for school districts that grew by 10 percent or more during the 1980 to 1990 period, the ratio approach tended to underestimate their total population by an average of 10.9 percent.

- One out of five school districts (21.1 percent) is located in areas where the population declined by 10 percent or more, representing 8.3 percent of the total population and 8.6 of the school-age population.

- Over one-third (36.5 percent) of the total population and 38.2 percent of the school-age population live in a district which had a major *increase* (10 percent or more) in population throughout the decade.[11]

- Overall, larger school districts tend to be growing whereas the smaller school districts appear to be declining in population.[12]

[Figure 4 about here]

Percent Poor School-age Children in 1980 (See Figure 5 and Tables 5 and 6)

- With the exception of school districts with zero percent poverty and 24 percent or more children in poverty, there is little difference in the weighted MAPEs.[13]

---

[11] Special tabulation by the U.S. Census Bureau.

[12] Special tabulation by the U.S. Census Bureau.

[13] When there were no related poor school-age children in 1980, then the shares methodology predicted that the percentage of children in poverty in 1990 will be zero as well.

- The unweighted MAPE in areas with a poverty rate of 24 percent or more is 16.5 percent for the total population and 21.3 percent for the school-age population.

- The bias in the estimates in school districts with a poverty rate of 24 percent or more approximates 9.3 percent for the total population and 8.4 percent for the school-age population.
- Approximately 16.5 percent of all school districts and 20.9 percent of all school-age children are in areas where the poverty rate is 24 percent or more.

[Figure 5 about here]

Percent Poor School-age Children in 1990 (See Figure 6 and Tables 5 and 6)

- The MAPEs for the percent poor school-age children in 1990 are similar to the MAPEs for the percent poor school-age children in 1980.[14]

- For school districts that have a poverty rate of 24 percent or more in 1990, we tend to overestimate both the total and school-age populations (MALPEs of 10.1 percent and 8.6 percent, respectively).

[Figure 6 about here]

Numerical Change in Poverty Rate for Children, 1980-1990 (See Figure 7 and Tables 5 and 6)

- The pattern for the MAPEs and MALPEs are similar to the MAPEs and MALPEs for the demographic characteristic representing population growth between 1980-1990.

- Like the growth rate for the population, there are higher MAPEs and MALPEs for both the total and school-age populations in school districts with large increases in the poverty rate and school districts with large decreases in the poverty rate.

Obviously, these situations occur in very, very small school districts. As a result, the predictions are not accurate and there is a high degree of error between the predictions and the truth.

[14] When there are no children in poverty, the percent difference (for that school district) is undefined and excluded from our tabulations. Even with the missing values removed, smaller school districts continue to contribute disproportionately to the high MAPEs.

- Nearly 97 percent of the school districts which experienced a decline in poverty of 10 percent or more were located in school districts with a population size of 5,000 or less.[15]

- Eighty percent of the school districts which experienced an increase in poverty of 10 percent or more were located in school districts with less than 20,000 people.[16]

[Figure 7 about here]

Census Division (See Figure 8 and Tables 5 and 6)

- The largest unweighted MAPE for both the total and school-age population is in the Mountain region. When weighted, the MAPE for the Mountain Region is aligned with the remaining census regions.

[Figure 8 about here]

Percent Hispanic in 1980 (See Figure 9 and Tables 5 and 6)

- MAPEs for the Hispanic population may be correlated with the percent of the population which is Hispanic.

[Figure 9 about here]

Percent Black in 1980 (See Figure 10 and Tables 5 and 6)

- The MAPEs for the total and school-age population are monotonic.

[Figure 10 about here]

Percent Group Quarters in 1980 (See Figure 11 and Tables 5 and 6)

- The MAPEs are higher for school districts where the GQ population comprises more than 10 percent of the total population. However, the percentage of GQ population is difficult to estimate over time because GQ facilities are built or closed over the decade.

[Figure 11 about here]

---

[15] Special tabulation by the U.S. Census Bureau.

[16] Special tabulation by the U.S. Census Bureau.

V. CONCLUSIONS AND PLANS TO IMPROVE SCHOOL DISTRICT ESTIMATES

This paper attempted to evaluate the 1990 school district level population estimates which were developed by the synthetic ratio approach. For both the total population and the school-age population age 5-17, four sets of synthetic estimates were produced: (1) the 1990 county estimates-based estimates; (2) the 1990 county count-based estimates; (3) state growth-based estimates; and (4) national growth-based estimates. To evaluate the estimates, we used both the Mean Absolute Percent Error (MAPE) and the Mean Algebraic Percent Error (MALPE). We examined the variations in the MAPEs and MALPEs by selected demographic and economic characteristics.

To summarize, the state growth and national growth-based models produced the least accurate estimates. They are feasible alternatives, but the school district growth rate is least likely to be the same as the state's or the nation's. The county count-based and the county estimates-based models were close to each other although the former provided more accurate estimates than the latter. We found that the differences were especially apparent for small school districts, districts with high and low poverty rates; and districts with high and low growth rates. However, the county count-based estimates can be produced only at the census year. Therefore, if we must rely on the synthetic estimates, we do need to use the county estimates-based model.

<u>What are our plans to improve the school district estimates?</u>

The Census Bureau is required to produce school district level population estimates for SY 1995/1996 and every two years thereafter. For SY 1995/1996 and SY 1997/1998, the synthetic estimates were based on data from the 1990 census and updated county estimates thereafter. For SY 1999/2000, we will use the Census 2000 data.

However, for post 2000 school district estimates, we plan to conduct further research to improve the estimates. These research plans include:

- Examine the use of updated TIGER/MAF files to more adequately define school district boundaries and aggregate blocks to the school district level.

- Evaluate the geocoding of addresses extracted from IRS tax returns to the school district level.

- Expand the use of administrative records, such as the extract of IRS tax returns, Common Core of Data, or Free Lunch data to directly and indirectly estimate total and school-age population estimates.

- Explore the use of the American Community Survey (ACS) data to estimate school district population.

REFERENCES

National Research Council, 1998. *Small-Area Estimates of Children in Poverty, Interim Report 2, Evaluation of Revised 1993 County Estimates for Title I Allocations*. Panel on Estimates of Poverty for Small Geographic Areas, C.F. Citro, M.L. Cohen, and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: National Academy of Press.

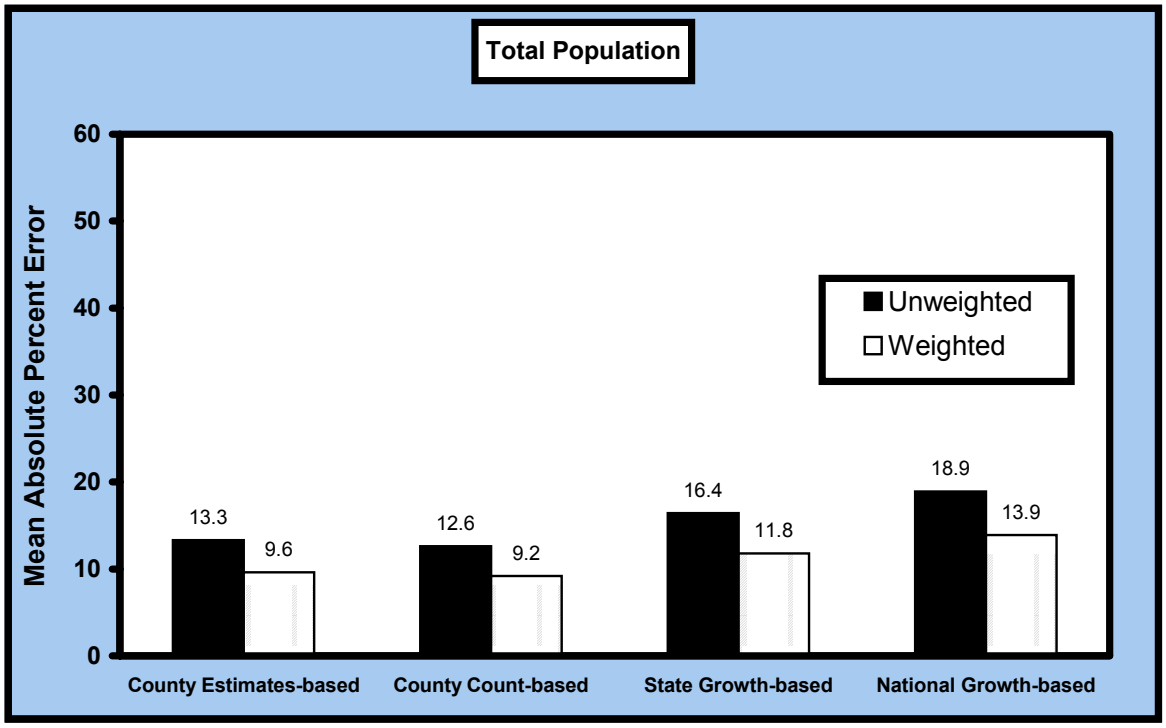
U.S. Census Bureau, 1997. *Table presented to the National Academy of Sciences, Panel on Estimates of Poverty for Small Geographic Areas*, Sixth Plenary Meeting, November 4, 1997.

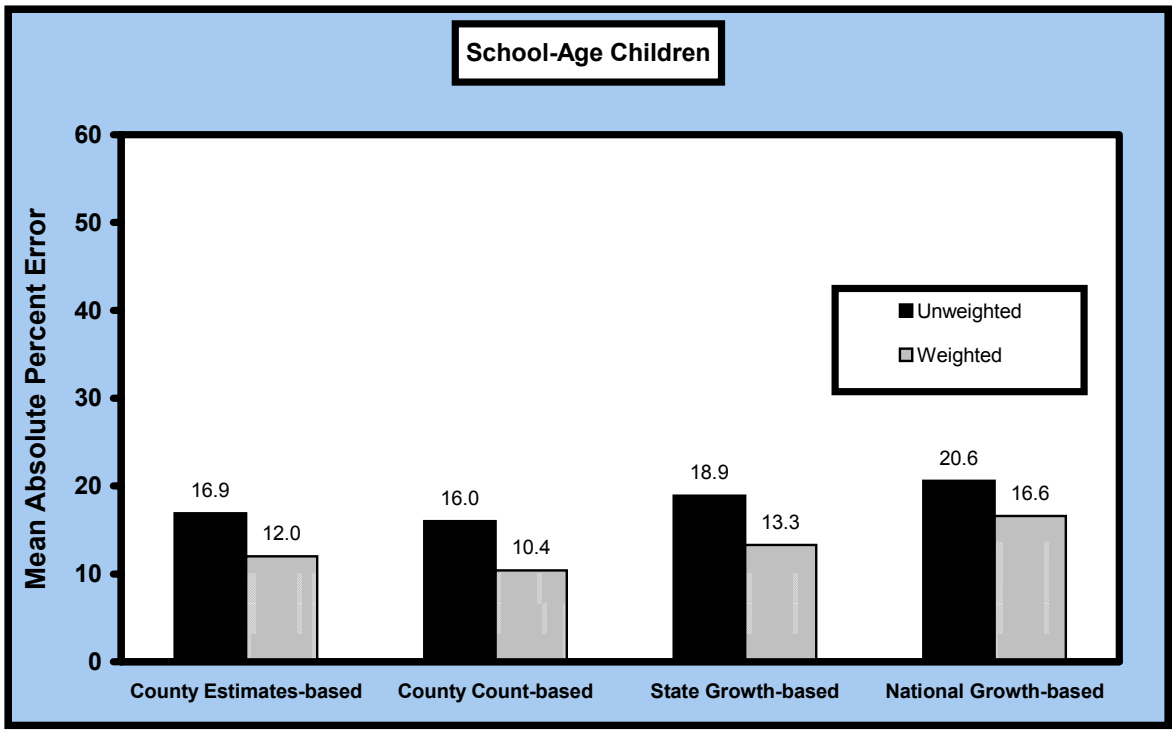
U.S. Census Bureau, 1998. *Appendix provided to the National Academy of Sciences, Panel on Estimates of Poverty for Small Geographic Areas*, Ninth Plenary Meeting, October 2-3, 1998.

U.S. Department of Education, 1999. *National Center for Education Statistics. Classification Evaluation of the 1994-95 Common core of Data: Public/Elementary/Secondary Education Agency Universe Survey*, NCES 1999-316 by Stephen Owens. Project Officer: Beth Young. Washington, DC.

Figure 1

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Method to Estimate 1990 School Districts

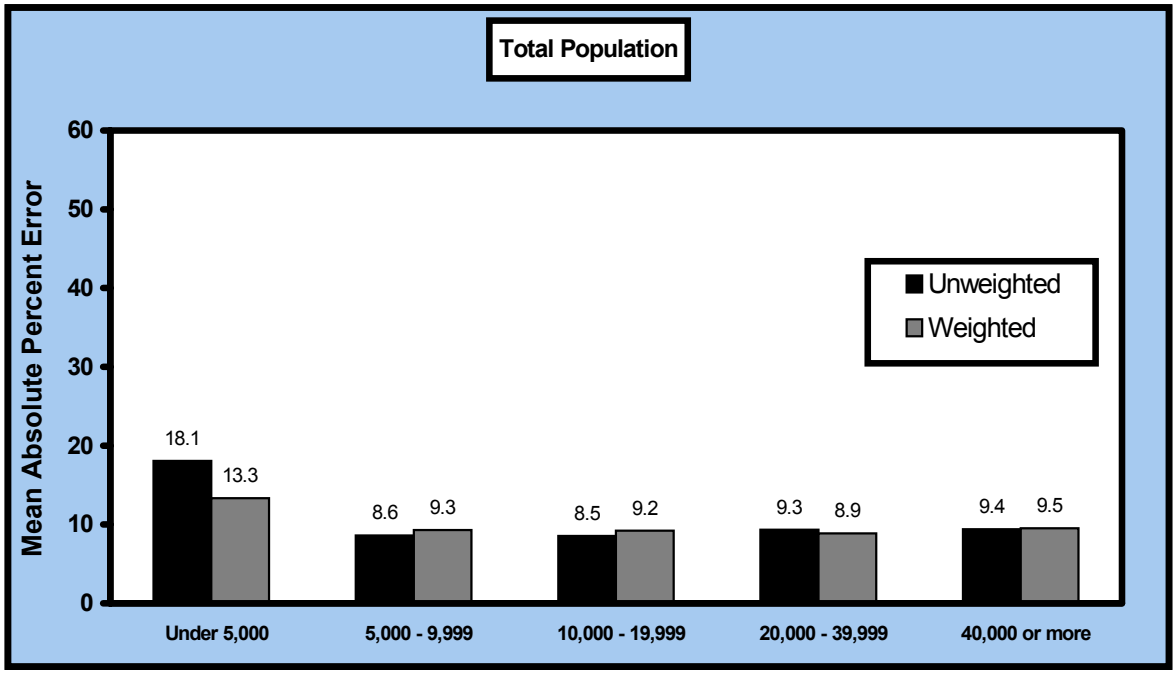


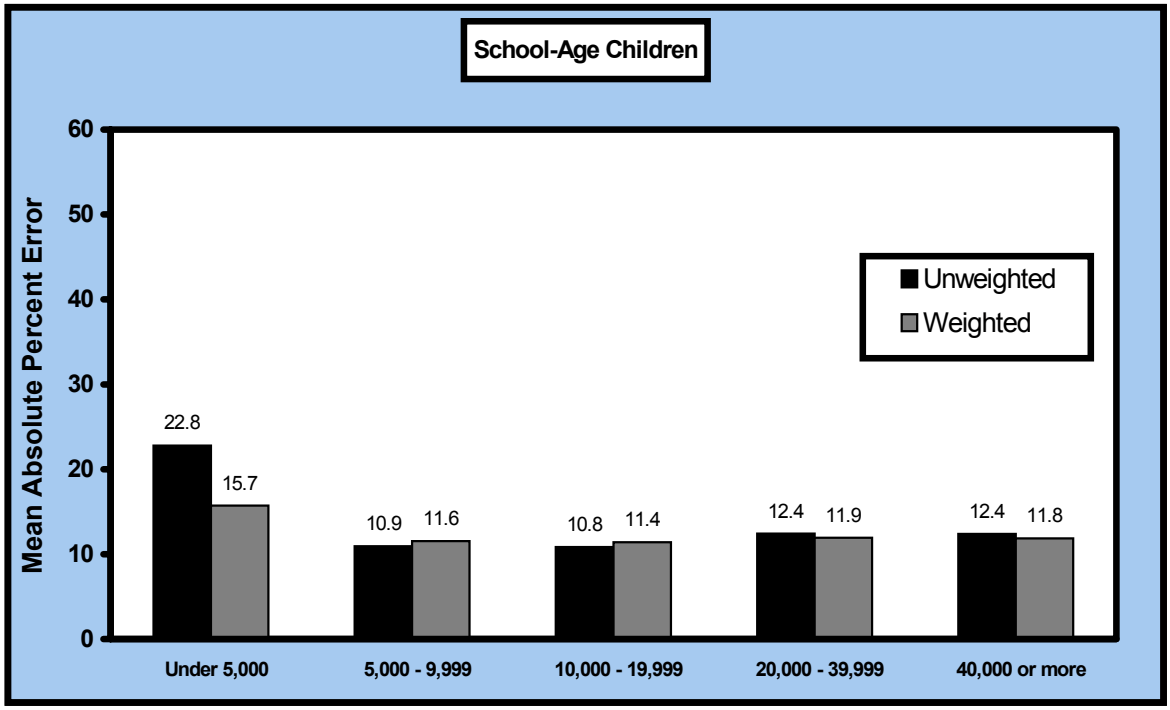


Source: U.S. Census Bureau

Figure 2

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Size of School District: 1980

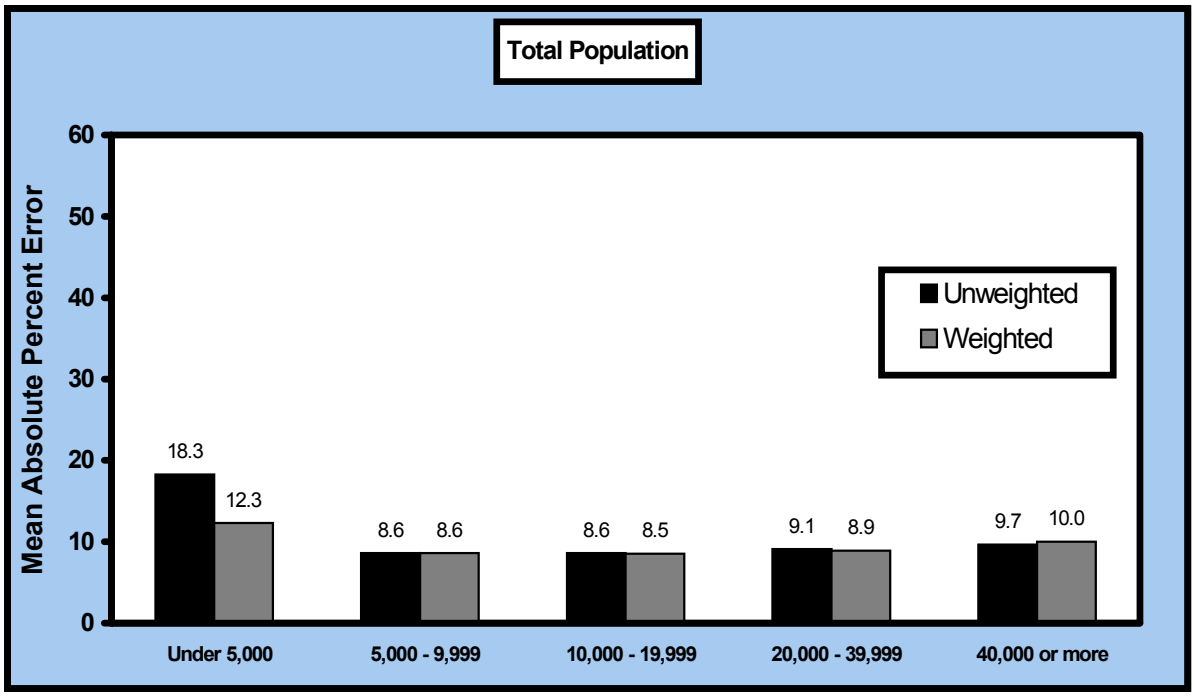


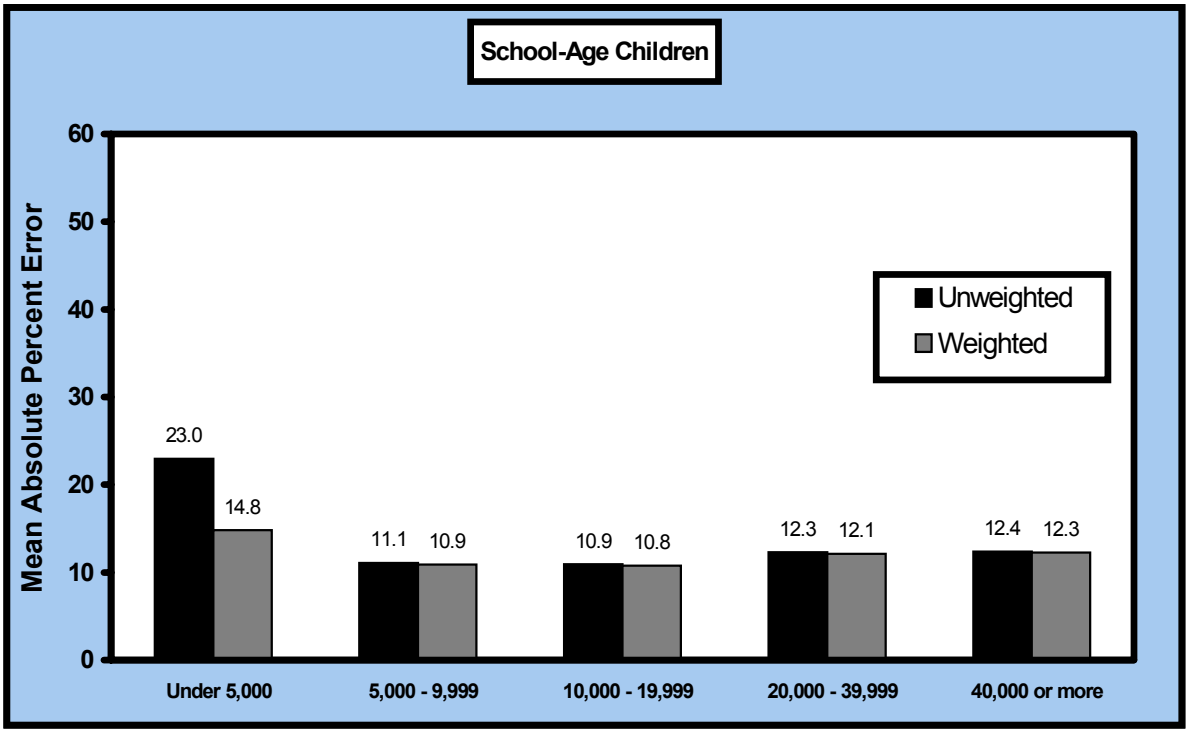


Source: U.S. Census Bureau

Figure 3

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Size of School District: 1990

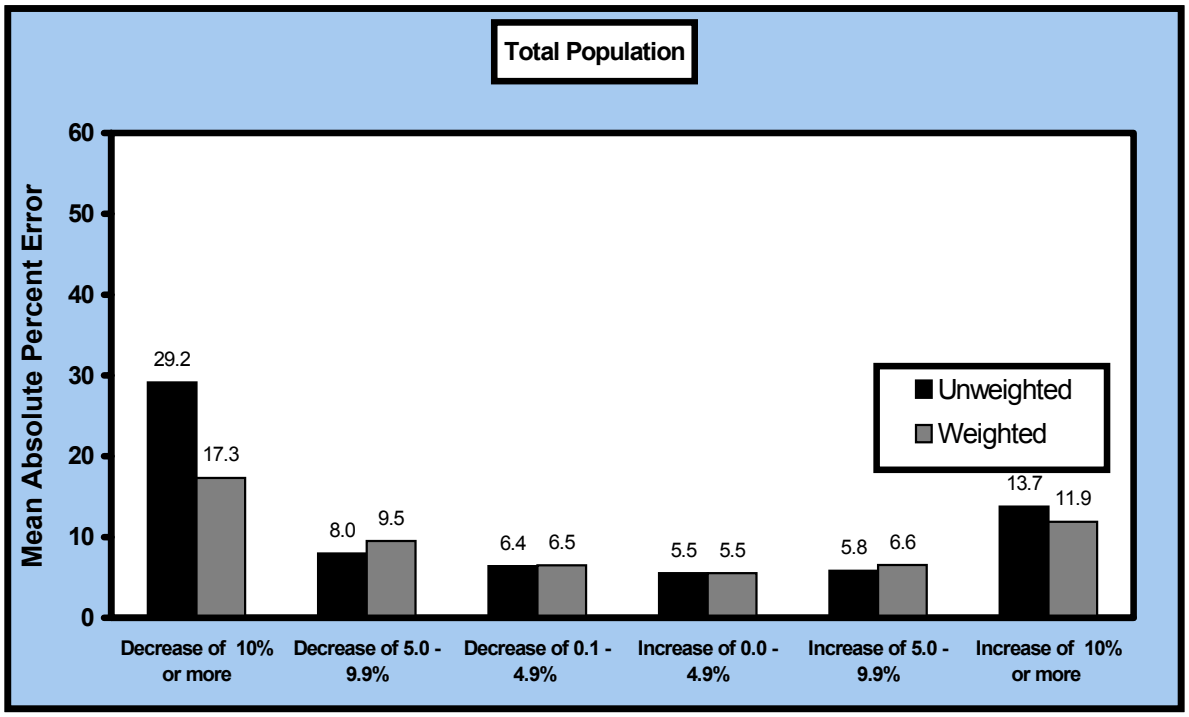


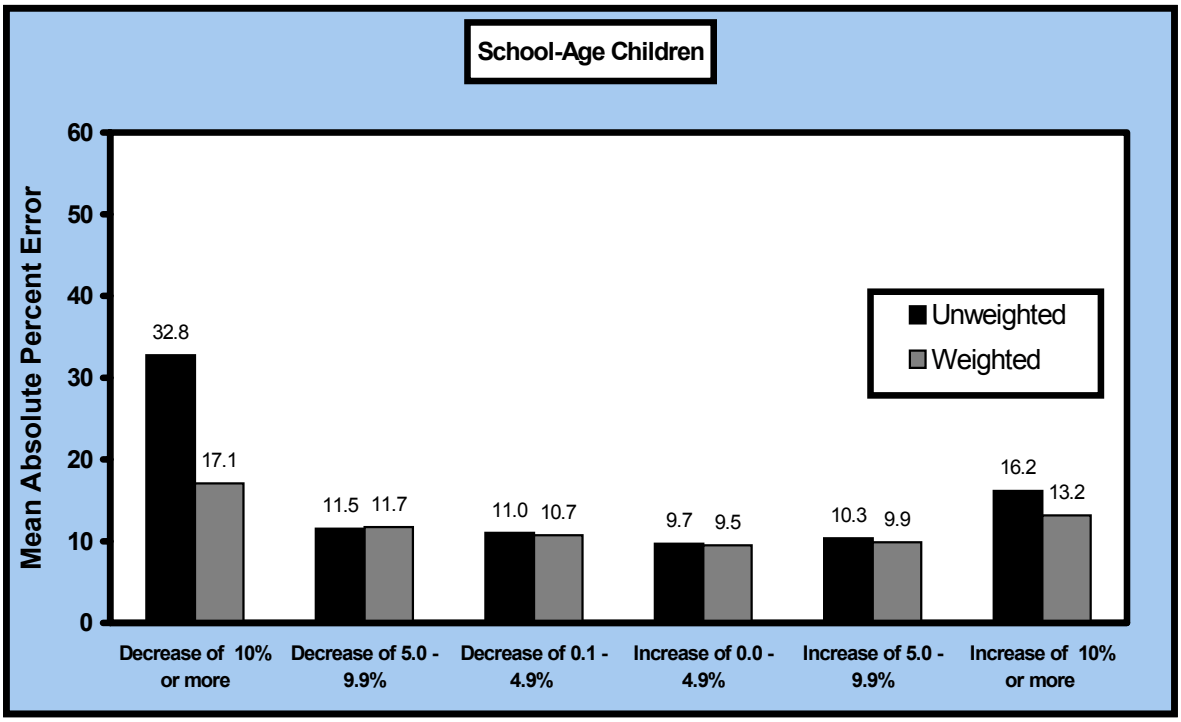


Source: U.S. Census Bureau

Figure 4

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Population Growth: 1980-1990

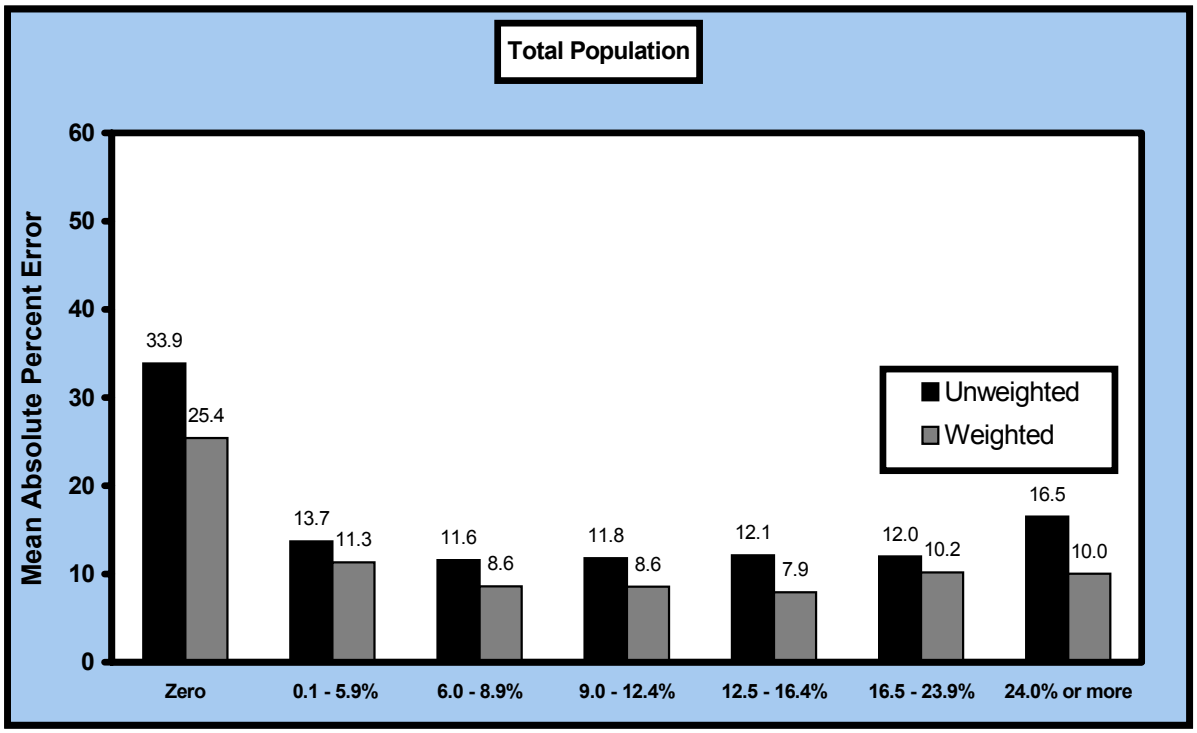


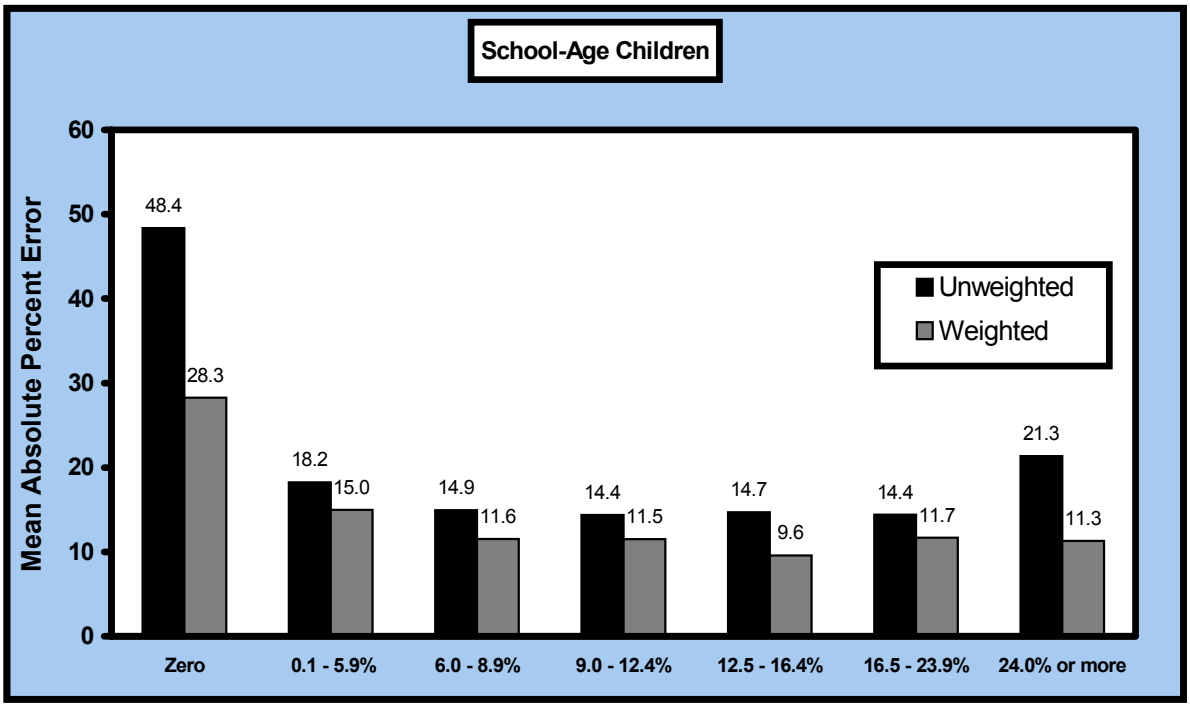


Source: U.S. Census Bureau

Figure 5

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Percent Poor School-age Children: 1980

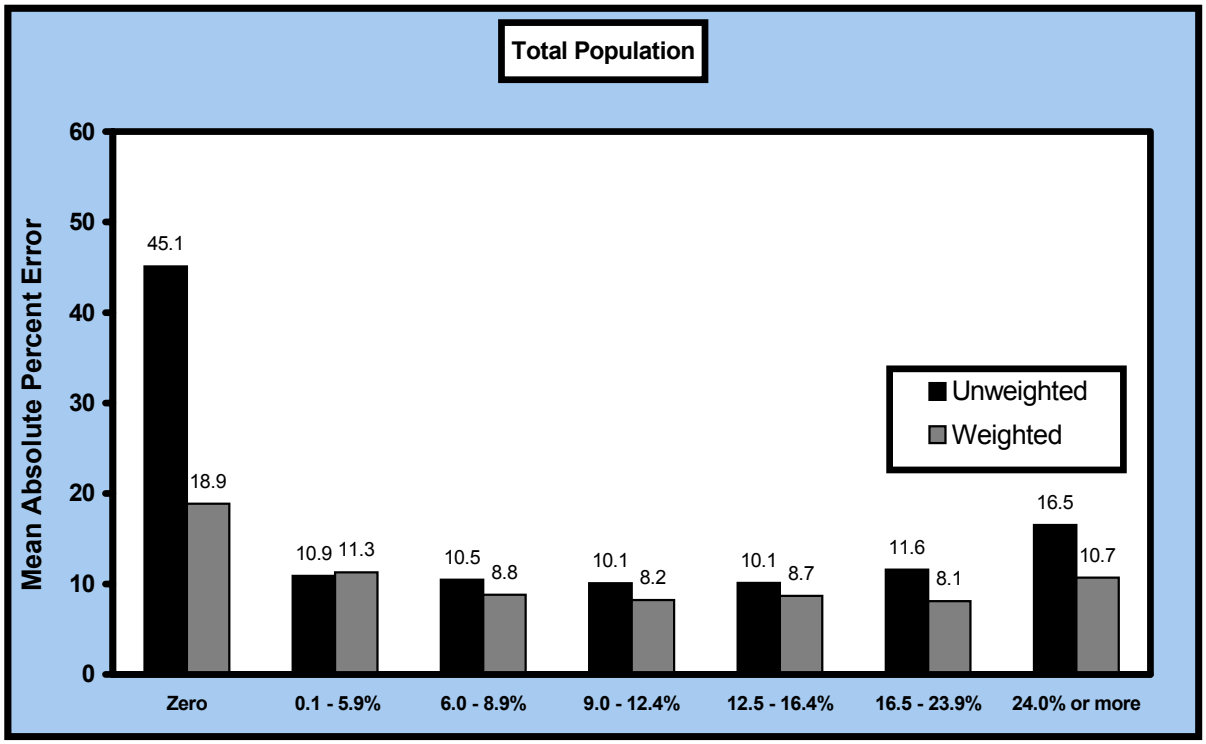


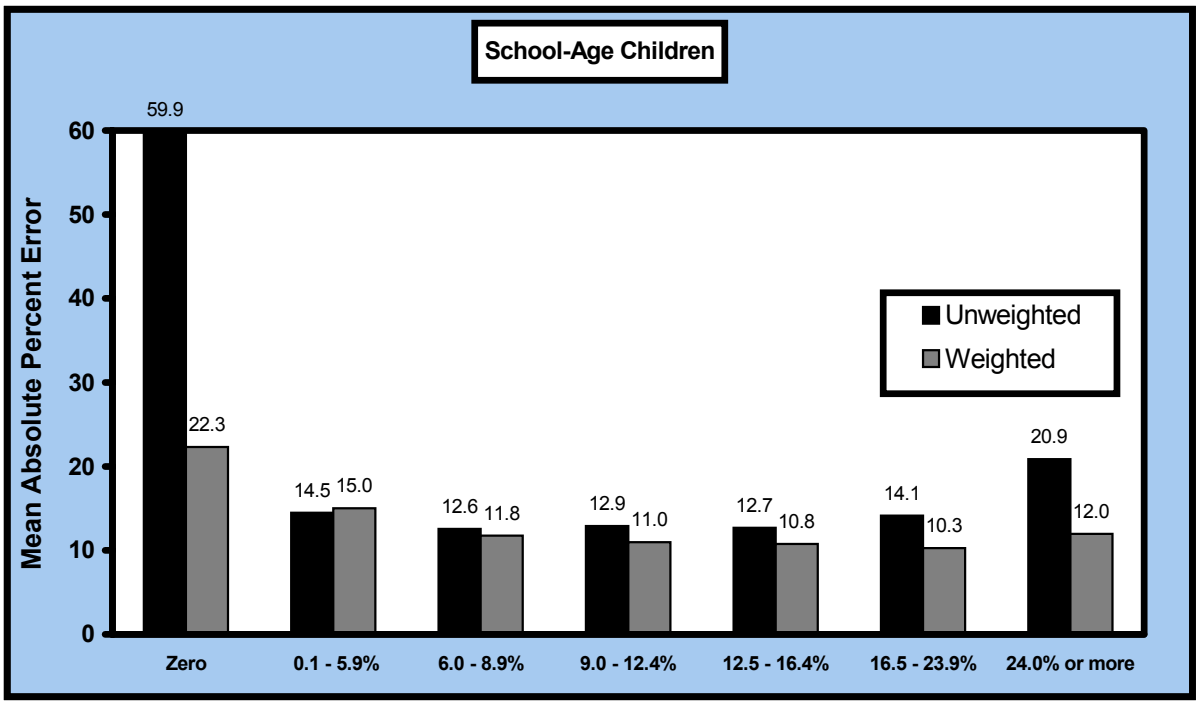


Source: U.S. Census Bureau

Figure 6

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Percent Poor School-age Children: 1990

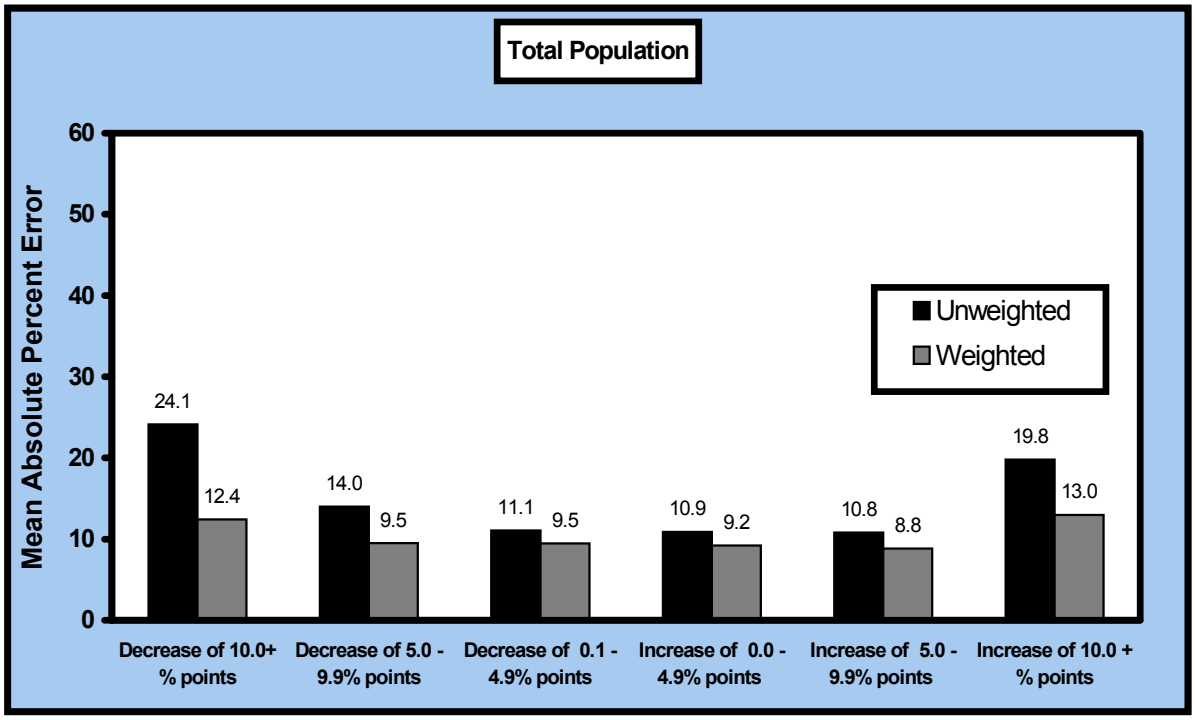


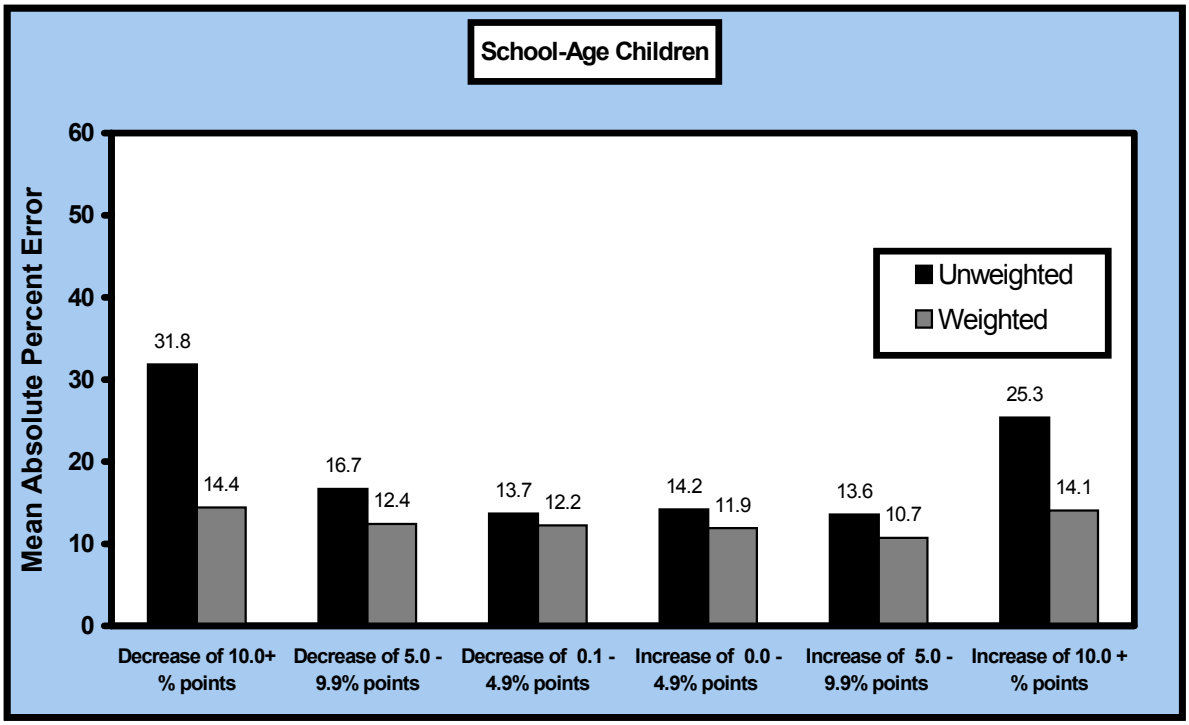


Source: U.S. Census Bureau

Figure 7

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Change in Poverty Rate for Children: 1980-1990

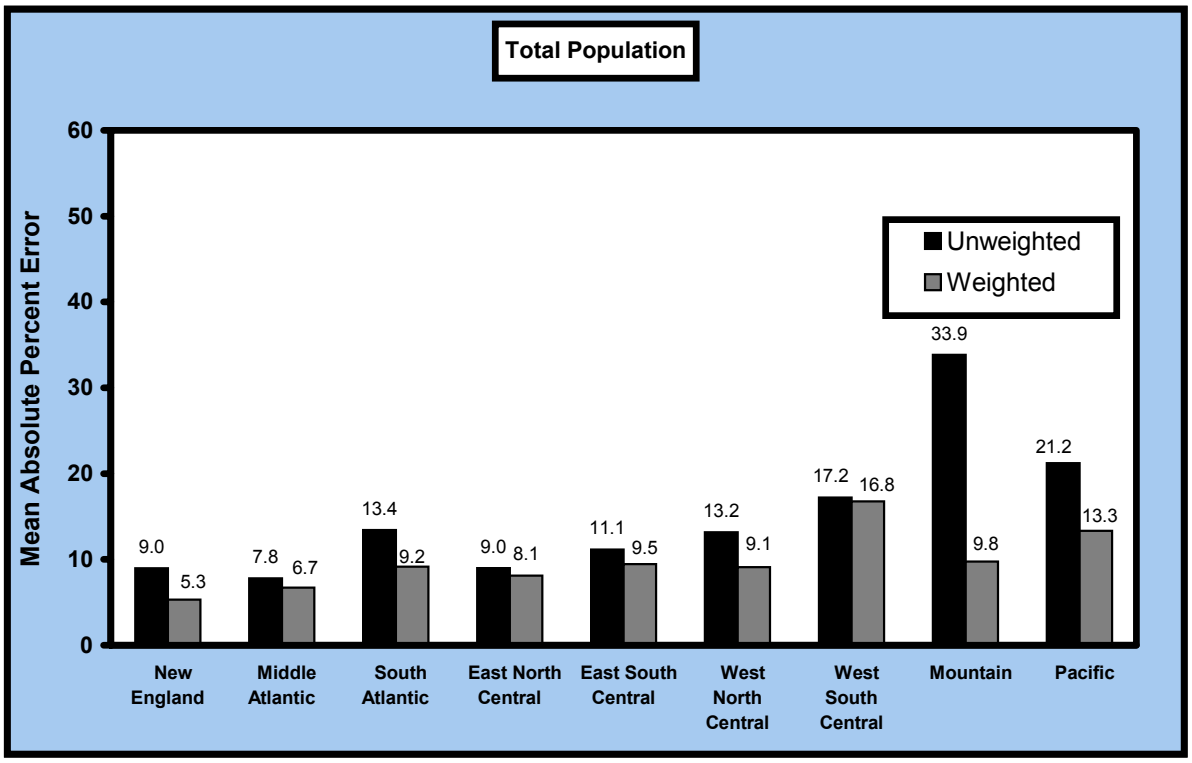


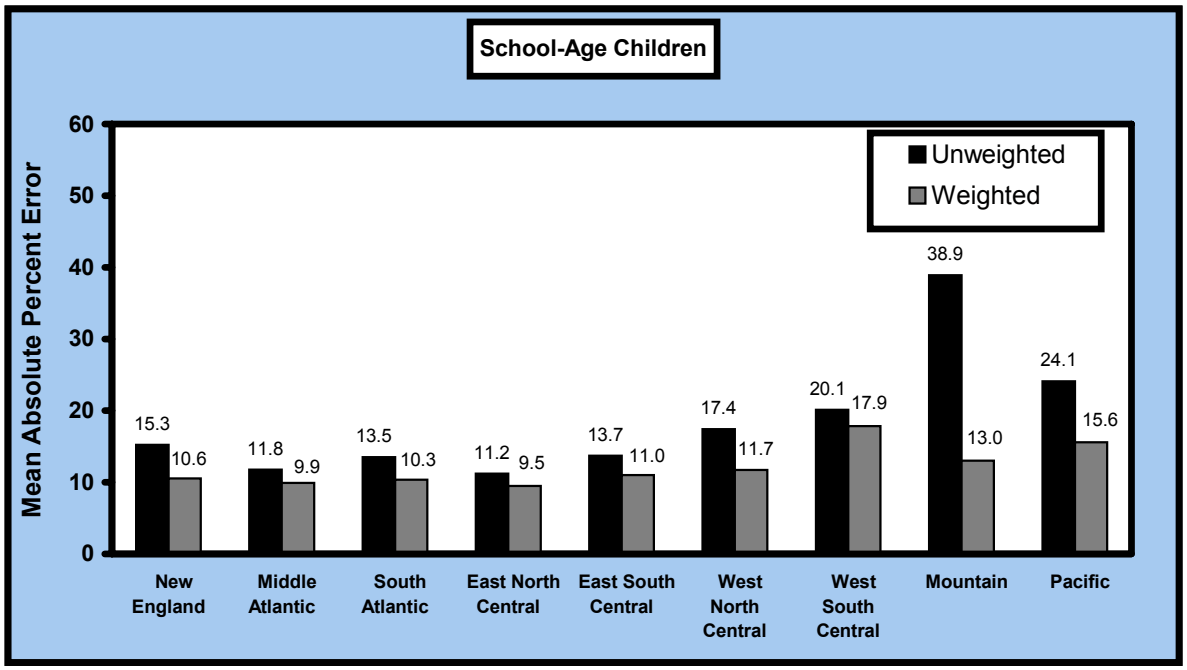


Source: U.S. Census Bureau

Figure 8

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Census Division: 1980

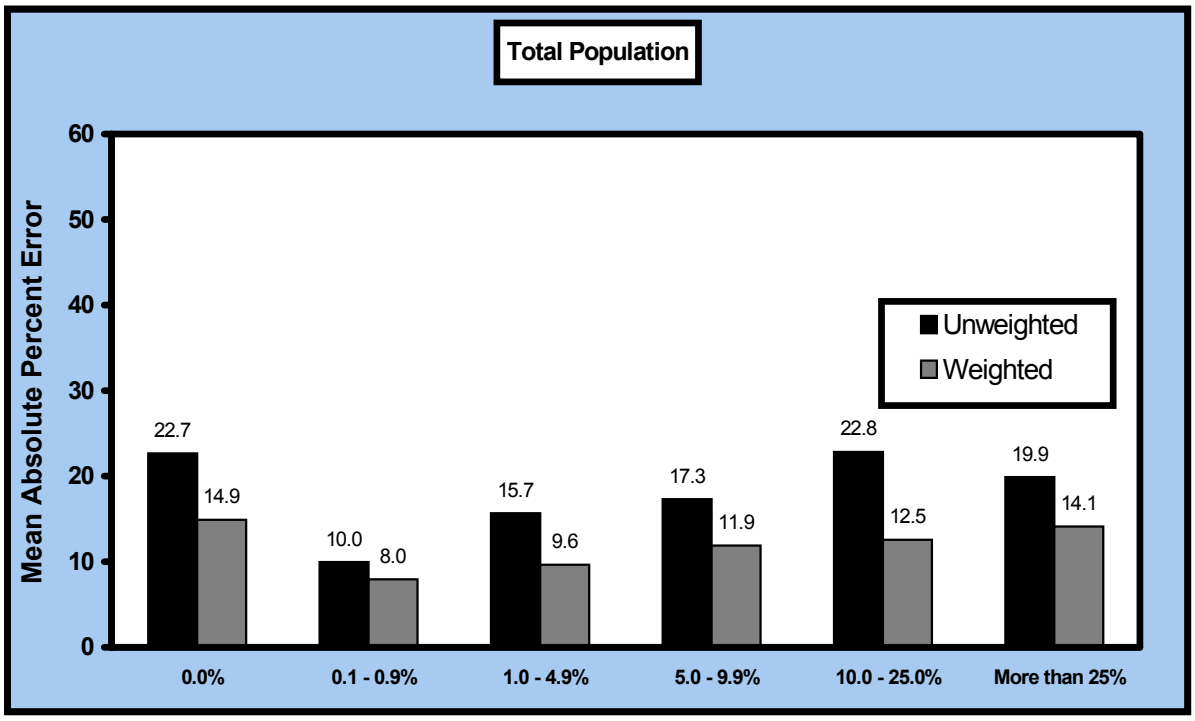


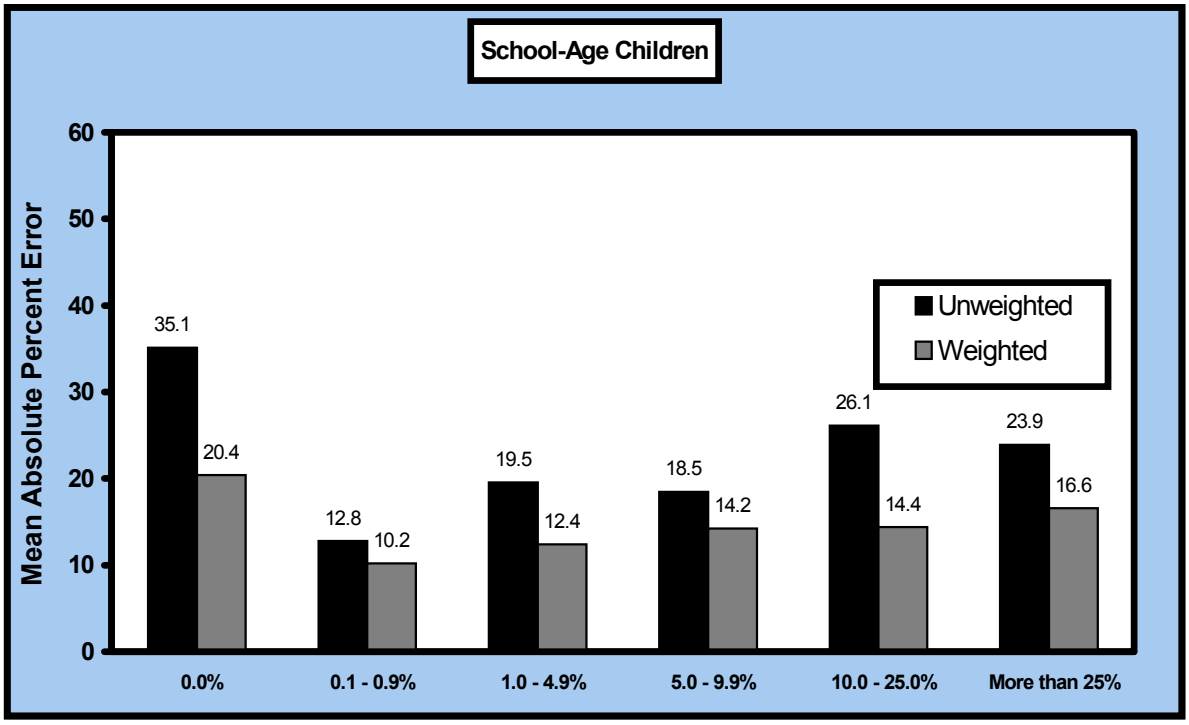


Source: U.S. Census Bureau

Figure 9

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Percent Hispanic: 1980

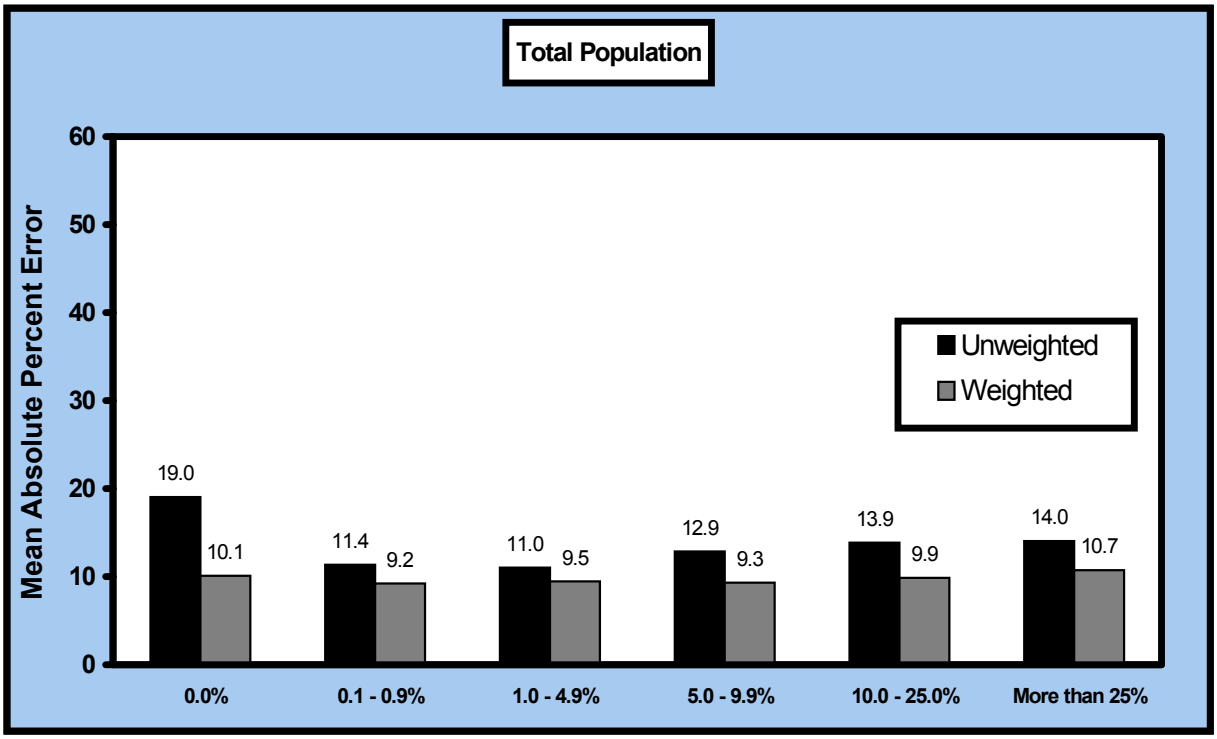


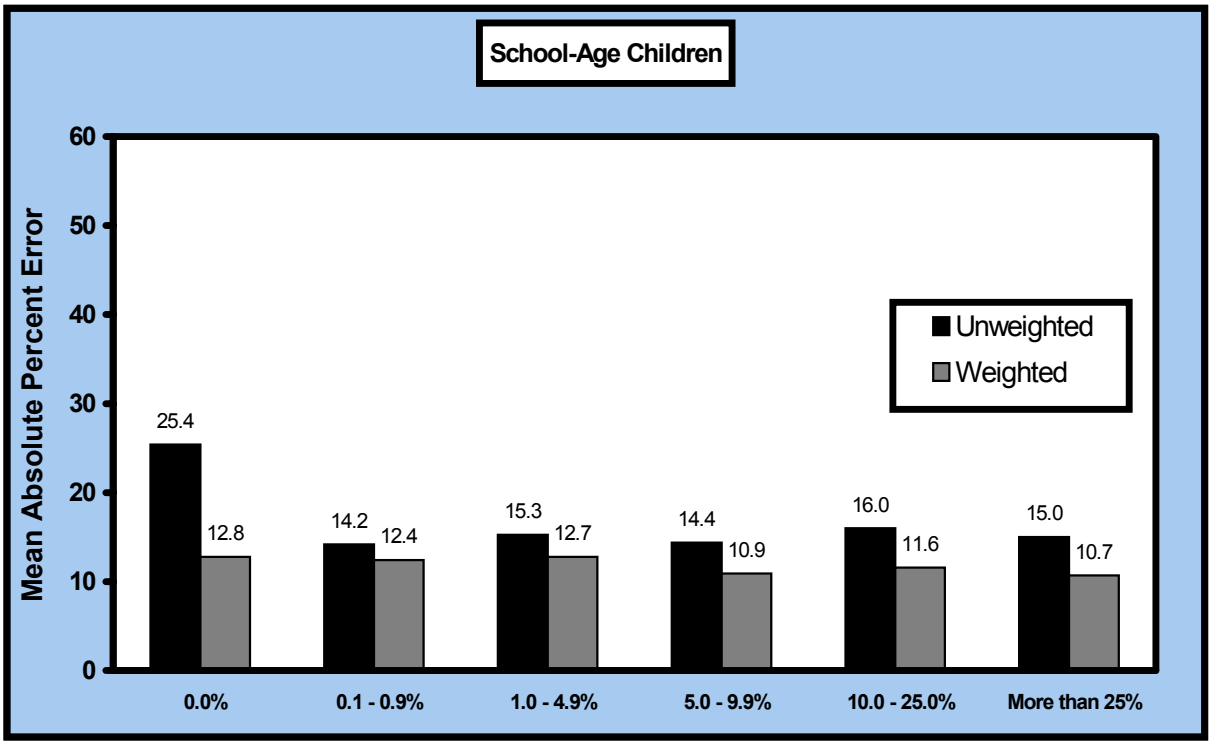


Source: U.S. Census Bureau

Figure 10

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Percent Black: 1980

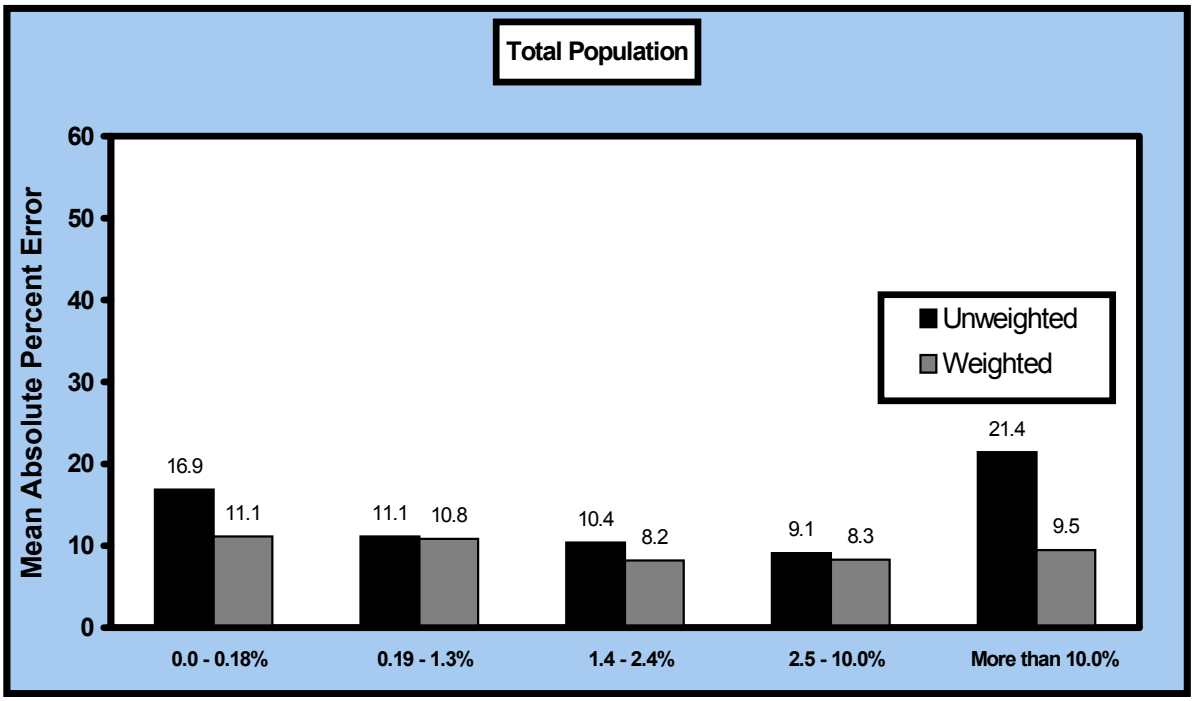


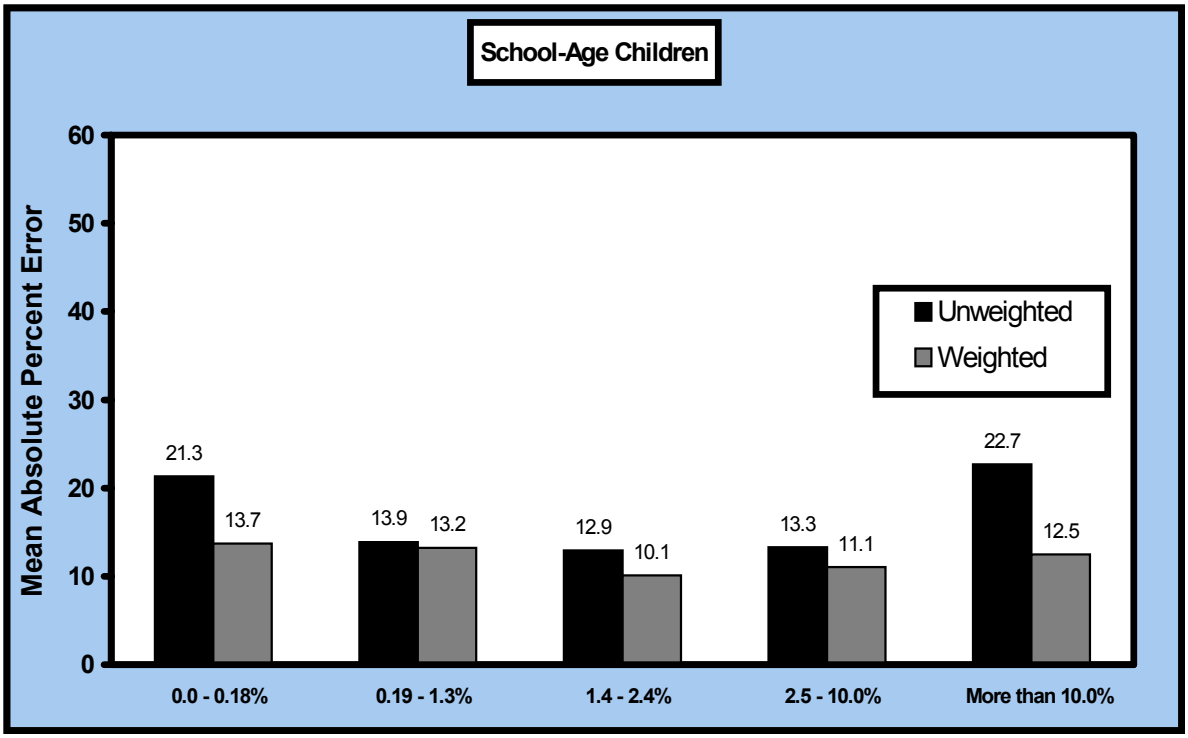


Source: U.S. Census Bureau

Figure 11

Mean Absolute Percent Errors (MAPEs) for Estimates of Total Population and School-age Population by Percent Group Quarters (GQ): 1980





Source: U.S. Census Bureau

August 15,2001

POPULATION DIVISION WORKING PAPER SERIES

For copies of these working papers, please contact the Statistical Information Staff, Population Division, Bureau of the Census, Washington, DC 20233-8800 (301)457-2422/pop@census.gov

NO. 1 - "The Census Bureau Approach for Allocating International Migration to States, Counties, and Places: 1981-1991." David L. Word. October 1992.

NO. 2 - "Geographic Coding of Administrative Records-- Past Experience and Current Research." Douglas K.Sater. April 1993.

NO. 3 - "Postcensal Population Estimates: States, Counties, and Places." John F. Long. August 1993.

NO. 4 - "Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census Post Enumeration Survey Results." R. Colby Perkins. August 1993.

NO. 5 - "Evaluation of Postcensal County Estimates for the 1980s." Sam T. Davis. March 1994.

NO. 6 - "Metropolitan Growth and Expansion in the 1980s." Richard L. Forstall and James D. Fitzsimmons. April 1993.

NO. 7 - "Geographic Coding of Administrative Records -- Current Research in ZIP/Sector-to-County Coding Process." Douglas K. Sater. June 1994.

NO. 8 - "Illustrative Ranges of the Distribution of Undocumented Immigrants by State." Edward W. Fernandez and J. Gregory Robinson. October 1994.

NO. 9 - "Estimates of Emigration of the Foreign-Born Population: 1980-l990." Bashir Ahmed and J. Gregory Robinson. December 1994.

NO. 10 - "Estimation of the Annual Emigration of U.S. Born Persons by Using Foreign Censuses and Selected Administrative Data: Circa 1980." Edward W. Fernandez. January 1995.

NO. ll - "Using Analytic Techniques to Evaluate the 1990 Census Coverage of Young Hispanics." Edward Fernandez. May 1995.

NO. 12 - "Metropolitan and Nonmetropolitan Areas: New Approaches to Geographical Definition." Donald C. Dahmann and James D. Fitzsimmons. October 1995.

NO. 13 - "Building a Spanish Surname List for the 1990's--
A New Approach to An Old Problem."
David L. Word and R. Colby Perkins, Jr.
February 1996.

NO. 14 - "Fertility of American Men."
Amara Bachu. March 1996.

NO. 15 - "Comparisons of Selected Social and Economic Characteristics Between Asians, Hawaiians, Pacific Islanders, and American Indians (Including Alaskan Natives)." Edward W. Fernandez. June 1996.

NO. 16 - "Findings on Questions on Race and Hispanic Origin Tested in the 1996 National Content Survey."
Prepared in the Population Division by the Racial Statistics Branch and the Ethnic and Hispanic Statistics Branch. December 1996.

NO. 17 - "Race and Ethnicity Classification Consistency Between the Census Bureau and the National Center for Health Statistics." Larry Sink. February 1997.

NO. 18 - "Results of the 1996 Race and Ethnic Targeted Test."
May 1997.

NO. 19 - "Who Responds/Who Doesn't? Analyzing Variation in Mail Response Rates During the 1990 Census." David L. Word.
July 1997.

No. 20 - "Trends in Marital Status of U.S. Women at First Birth: 1930 to 1994." Amara Bachu. March 1998.

No. 21 - "State Estimates of Organized Child Care Facilities."
Lynne Casper and Martin O'Connell. March 1998.

No. 22 - "How Well Does the Current Population Survey Measure the Foreign-Born Population in the United States"
Dianne Schmidley and J. Gregory Robinson. April 1998

No. 23 - "Poverty, Family Structure, and Child Well-Being: Indicators from the SIPP." Jason Fields and Kristin Smith. April 1998

No. 24 - "Child Well-Being Indicators from the SIPP."
Kristin Smith, Loretta Bass, and Jason Fields.
April 1998

No. 25 - "Timing of First Births: 1930-34, 1990-94."
Amara Bachu. May 1998

No. 26 - "Co-Resident Grandparents and Grandchildren: Grandparent Maintained Families."
Lynne Casper and Ken Bryson
May 1998

No. 27 - "Population of the 100 Largest Cities and Other Urban Places in the United States: 1790 to 1990." Campbell Gibson. June 1998

No. 28 - "Are There Differences in Voting Behavior Between Naturalized and Native-born Americans?" Loretta E. Bass and Lynne M. Casper. March 1999.

No. 29 - "Historical Census Statistics on the Foreign-born Population of the United States: 1850-1990." Campbell J. Gibson. February 1999

No. 30 - "Direct Measures of Poverty as Indicators of Economic Need: Evidence from the Survey of Income and Program Participation." Kurt J. Bauman. November 1998

No. 31 - "American Community Survey and Intercensal Population Estimates: Where Are the Cross-roads?" Amy Symens. December 1998

No. 32 - "Women's Labor Force Attachment Patterns and Maternity Leave: A Review of the Literature." Kristen Smith and Amara Bachu. January 1999

No. 33 - "Evaluation of Relationship, Marital Status, and Grandparents Items on the Census 2000 Dress Rehearsal" Charles Clark and Jason Fields. April 1999.

No. 34 - "Unbinding the Ties: Edit Effects of Marital Status on Same Gender Couples." Charles Clark and Jason Fields. April 1999.

No. 35 - "Racial-Ethnic and Gender Differences in Returns to Cohabitation and Marriage: Evidence from the Current Population Survey." Philip N. Cohen. May 1999.

No. 36 - "How Does POSSLQ Measure Up? Historical Estimates of Cohabitation." Lynne Casper, Philip N. Cohen, and Tavia Simmons. May 1999.

No. 37 - "Childlessness Among American Women On the Rise?". Amara Bachu. May 1999.

No. 38 - "Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100" Frederick Hollman, Tammany Mulder, and Jeffrey Kallan. October 1999.

No. 39 - "What Do We Know About the Undercount of Children?" Kirsten K. West and J. Gregory Robinson. August 1999.

No. 40 - Canceled

No. 41 - Canceled

No. 42 - "Measures of Help Available to Households in Need: Their Relationship to Well-Being, Welfare, and Work."

Kurt Bauman and Barbara Downs. May 2000

No. 43 - "Have We Reached the Top? Educational Attainment Projections of the U.S. Population." Jennifer Cheeseman Day and Kurt Bauman May 2000

No. 44 - "The Emerging American Voter: An Examination of the Increase in the Black Vote in November 1998." Avalaura L. Gaither and Eric C. Newburger. June 2000

No. 45 - "An Analysis of State and County Population Changes by Characteristics: 1990-1999." Amy Symens Smith, Bashir Ahmed, and Larry Sink. November 2000.

No. 46 - "The Effect of Work and Welfare on Living Conditions in Single Parent Households." Kurt Bauman. August 2000.

No. 47 - "Projections of the Number of Households and Families in the United States: 1999 to 2025." Monique Oosse and Ching li Wang. Release Pending.

No. 48 - "Fertility in the United States: Is State-Level Fertility Converging to the National Level?" Laura Heaton. Release Pending.

No. 49 – “Variation in State Morality over Time.” Monica Oosse. Release Pending.

No. 50 - “Accuracy of the U.S. Census Bureau National Population Projection and Their Respective Components of Change.” Tammany Mulder. Release Pending.

No. 51 - “U.S. Census Bureau Measurement of Net International Migration to the United States: 1990 to 2000.” Tammany Mulder, Frederick Hollmann, Lisa Lollock, Rachel Cassidy, Joseph Costanzo, and Josephine Baker. Release Pending.

No. 52 - “At-Risk Conditions of U.S. School-Age Children.” Robert Kominski, Amie Jamieson, and Gladys Martinez. Release Pending.

No. 53 - “Home Schooling in the United States: Trends and Characteristics.” K. J. Bauman. August 2001.

No. 54 - “Evaluation of the 1990 School District Level Population Estimates Based on the Synthetic Ratio Approach.” Esther R. Miller Release Pending.

No. 55 - “State Estimates of Child Care Establishments 1977-1997.” Grace O’Neil and Martin O’Connell