Methodology

Data for each indicator are drawn from publicly available sources. Each indicator used the most recently available data as of March 2014. Wherever possible, the *Index* uses annually updated data.

The change over a period of years is reported for each of the indicators. For most indicators, this report includes 10-year, five-year, and one-year changes. The only exceptions are those for which annual data are not available, and those exceptions are noted on their charts.

In addition, the comprehensive table on pages 78–79 reports results from regression analysis and their statistical significance for each indicator. Regression analysis uses all indicator data over the given time period to estimate, or "draw," a "best-fitting" straight line through the data points for those years. This enables us to get a sense of the general trend over time and smooth out year-to-year variations that might be true changes or artifacts of methodological or sampling issues.

For most indicators, the regression models include available data points from the past decade. Thus, the slopes of the linear regression lines through the data points for an estimated period present the general trends in indicators for those years. Beyond understanding this general trend, however, these regression models should not be used for statistical inference.

For all but one indicator, the comprehensive table presents simple linear regression models in which the years are the independent "X" variables and the indicators are the dependent "Y" variables. When looking at the past 10 annual changes in data, the slope of the linear regression line provides an estimate of the overall change per year.

For the indicator variable "Money Taxed Away by Federal Government," we used a first-order autoregressive model to capture the dynamic nature of revenues over time. In this instance, we developed a regression model aiming to predict 2014's (projected) revenues from 2013's, 2013's from 2012's, etc. Thus, this first-order autoregressive model is nothing but a simple linear regression using one year's revenues as the dependent variable and the prior year's revenues as the independent variable.

For three indicators—the percentage that attends religious services weekly, the percentage of sexually experienced 12th-graders, and charter school enrollment—only biannual data were available. Consequently, the annual average change was determined using the past five rather than 10 changes in those cases.

One indicator—the reading proficiency of 17-year-olds—is drawn from a data source available only for every four years.

Indicators included in this report are illustrative and not a direct or exhaustive measure of culture and opportunity.

Definitions

Rate: A rate indicates the actual occurrences of a certain event given the number of possible events or a population that could experience the event. Rates are calculated by the event number divided by the

possible number of events or the population studied. For example, the marriage rate per 1,000 is the annual number of marriages that took place divided by the population of unmarried women age 15 and older for that year, and then multiplied by 1,000.

Percentage: A percentage is calculated by dividing the number in a specific category by the entire category and then multiplying that ratio by 100. For example, the percent of individuals in poverty is calculated by the number of individuals whose incomes fall below the official federal poverty line divided by the total population and then multiplied by 100.

Regression Slope Coefficient: This illustrates the value of how much change there is in the dependent "Y" variable, or the indicator, for one unit of change in the "X" independent variable, or the year. In other words, the slope coefficient estimates the annual change for the period estimated. A negative number indicates a decreasing trend, and a positive number, an increasing trend. For example, the 10-year (2001 to 2011) regression slope coefficient for the marriage rate is –1.03, meaning that during this period, each year, on average, the marriage rate decreased by 1.03 marriages per 1,000 unmarried women age 15 and older.

Regression Intercept Coefficient: This is the point where a regression line crosses the vertical or Y-axis—that is, the value of Y when X is zero. The intercept is often difficult to interpret meaningfully, depending on the regression model analyzed, but is included in the comprehensive table as it is the convention to report this regression estimator.

Statistical Significance and Goodness of Fit of Regression Coefficients: For each of our 36 indicators, we estimated a simple linear regression model using time as the independent variable and the indicator as the dependent variable, testing the null hypothesis that the regression slope coefficient was zero against a two-sided alternative. The p-value is defined as the probability of obtaining as extreme a test statistic (for the pertinent slope coefficient) as what was actually observed under the assumption that the null hypothesis is true. A slope coefficient estimate is typically determined to be statistically significant if its p-value fell below the critical threshold of α=0.05. We also similarly tested our intercept coefficients.

The coefficient of determination, R^2, provides a diagnostic regarding the goodness of fit of the regression model.