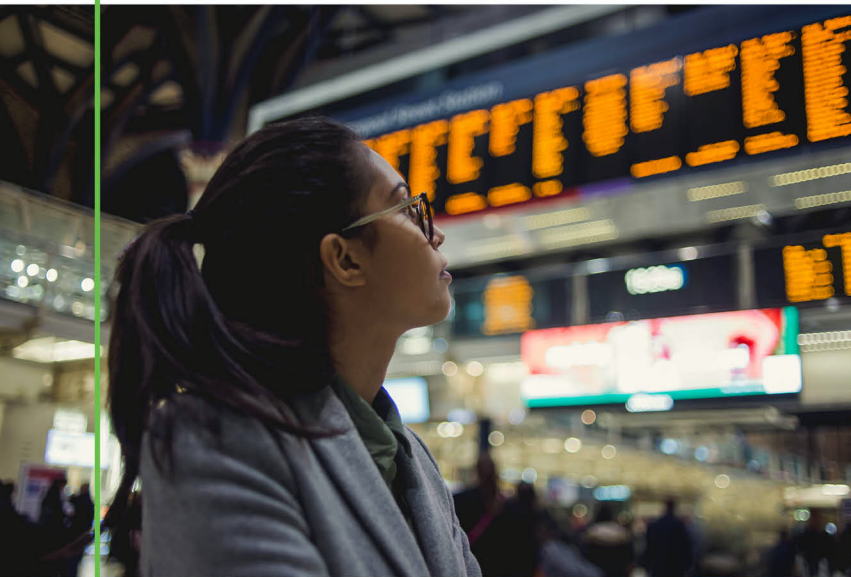# Assessing the Economic Gains of Eradicating Illiteracy Nationally and Regionally in the United States

JONATHAN ROTHWELL, PH.D.

Principal Economist, Gallup
Nonresident Senior Fellow, Brookings Institution, Metropolitan Policy Program

Copyright Standards

This document contains proprietary research, copyrighted materials and literary property of Gallup, Inc. It is for the guidance of your organization only and is not to be copied, quoted, published or divulged to others outside your organization. All of Gallup, Inc.'s content is protected by copyright. Neither the client nor the participants shall copy, modify, resell, reuse or distribute the program materials beyond the scope of what is agreed upon in writing by Gallup, Inc. Any violation of this Agreement shall be considered a breach of contract and misuse of Gallup, Inc.'s intellectual property.

This document is of great value to Gallup, Inc. Accordingly, international and domestic laws and penalties guaranteeing patent, copyright, trademark and trade secret protection safeguard the ideas, concepts and recommendations related within this document.

No changes may be made to this document without the express written permission of Gallup, Inc.

Gallup® is a trademark of Gallup, Inc. All rights reserved. All other trademarks and copyrights are property of their respective owners.

Summary

At the individual level, it is well-established that literacy is linked to a variety of positive outcomes, such as higher education, income, health and civil engagement.[1] Likewise, macroeconomic research has long-shown that higher levels of human capital — measured by educational attainment, cognitive ability and literacy — are associated with higher economic growth at the national level.[2]

Yet, according to a recent study from the Department of Education, roughly half of U.S. adults, aged 16 to 74 years old — 54% or 130 million people — lack literacy proficiency.

— ANALYSIS OF DATA FROM US. DEPARTMENT OF EDUCATION

Taking this figure into account, in conjunction with the macroeconomic evidence that human capital can drive economic growth, it follows that enormous economic gains would come from eradicating illiteracy in the U.S. By estimating the average income gains an individual could expect when moving from below-proficiency in literacy to the minimum proficiency, and then scaling those gains to the metropolitan, county, state and national levels, this study provides a simplified but plausible estimate of the potential impact of literacy gains in the U.S., from the individual to the national level.

These estimates are made possible by the release of new data from an international assessment of adult skills called the Program for the International Assessment of Adult Competencies (PIAAC) and the U.S. Department of Education. The Department of Education has recently combined individual PIAAC data from 2012 to 2017 to create and publish estimated literacy levels for every U.S. county.[3]

This document summarizes the findings and methods used to produce these estimates for the Barbara Bush Foundation and attempts to answer questions readers might have about how these estimates were produced.

Findings

Income is strongly related to literacy. The average annual income of adults who reach the minimum level for proficiency in literacy (Level 3) is nearly $63,000, significantly higher than the average of almost $48,000 earned by adults who score just below proficiency (Level 2) and much higher than those at low Levels of literacy (Levels 0 and 1), who earn just over $34,000 on average.

Mean Income in 2017 by Level of Literacy (2020 USD)

Source: 2012-2017 PIAAC U.S. State and County Estimates for Population; 2017 PIAAC for Income

■ Mean annual income, 2020 USD

Level	Mean annual income, 2020 USD
Levels 0-1	$34,127
Level 2	$47,596
Level 3	$62,997
Levels 4-5	$73,284
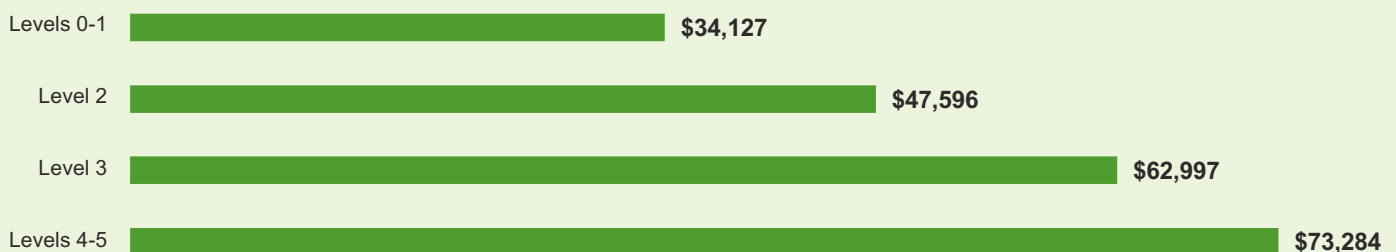
As with education, people with varying levels of literacy differ in other ways that also lead to income disparities, including age, gender, urbanicity, race, ethnicity, nativity, parental education and investment in learning during childhood. PIAAC includes measures of each of these. Controlling for these factors, the income-mean difference between people at or below Level 1 and those at Level 3 is $23,979 (instead of the raw gap of $28,870). The mean-income disparity between people at Level 2 and those at Level 3 is $13,193 (down from the unadjusted gap of $15,400).

These demographically adjusted income differences are used to estimate aggregate income gains for the nation and various geographies. They are preferred to the raw differences in average income because they hold constant characteristics of people that would remain the same if they were to gain literacy skills (like age, gender and race).

Eradicating illiteracy would have enormous economic benefits. This analysis finds that getting all U.S. adults to at least a Level 3 of literacy proficiency would generate an additional $2.2 trillion in annual income for the country. That is 10% of the gross domestic product.

There are several reasons why this could be considered a plausible but rough estimate of actual benefits. It is plausible because it scales up actual individual income differences between people with different levels of literacy, even after controlling for demographic characteristics. Thus, these estimated income gains from literacy hold demographic characteristics constant, including those of age, race, urbanicity and parental education, which can affect income regardless of literacy.

This should, nonetheless, be considered a rough estimate because such a large-scale change would have many indirect consequences. On the negative side, a massive increase in literacy would create some downward wage pressure on jobs that require high levels of literacy by increasing the supply of workers, and would create upward pressure on wages at the lower end of the literacy distribution. On the plus side, these supply effects would reduce income inequality, and an increase in literacy would likely boost entrepreneurship, productivity and innovation in large but unpredictable ways that benefit of the economy.

The areas with the lowest current levels of literacy would stand to gain the most as a share of the GDP.

This analysis shows the gains that would come from eradicating illiteracy for each state, county and metropolitan area, divided by the GDP to assess the proportional effect on each area's economy.

If states are ranked by their potential gains as a share of the GDP, four of the six states that stand to gain the most are in the South (Figure 1). This is largely because these states have a disproportionate share of adults with low levels of literacy proficiency. In Alabama, gains would be 15.6% of the GDP. An estimated 61% of Alabama's adult population falls below Level 3 on the PIAAC literacy scale and is therefore considered nonproficient in literacy. At the same time, the individual gains from education (and hence literacy) are roughly as large in Alabama as they are nationally, so adjusting the estimates to local conditions does not diminish the effect. Arkansas is in a similar position and would see estimated gains worth 14.4% of the GDP.

At the other end, gains from eradicating illiteracy would be relatively small in Washington, D.C. (5% of the GDP). A relatively low share of Washington, D.C.'s population is nonproficient (47%), even though the gains from education and literacy are high, given the favorable labor market for professionals. Meanwhile, in North Dakota, there are relatively high-paying opportunities for less educated workers, so the individual gains from literacy are comparably smaller (roughly 60% of the national gains). Moreover, North Dakota also has low rates of nonproficiency (45%). Taken together, this explains why North Dakota would only see income gains worth 3.9% of the GDP.

Figure 1. Gains From Eradicating Illiteracy as a Share of GDP
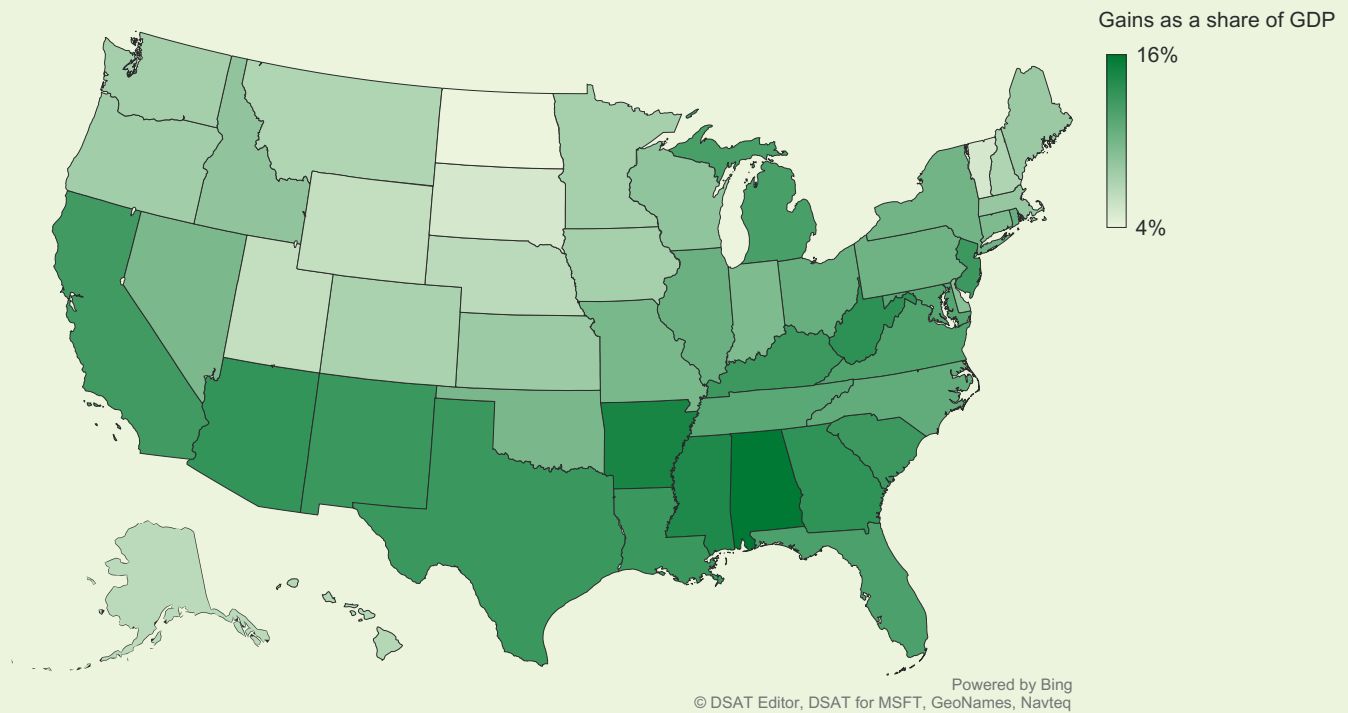
Looking across metropolitan areas, places with large Latin American immigrant populations — like McAllen, Texas, and Merced, California — stand to gain the most from eradicating illiteracy. In McAllen, an estimated 84% of the population would score below Level-3 proficiency and would gain an estimated 41% of the GDP if they became proficient. Merced would gain an estimated 26% of the GDP, largely because 72% of its adults are nonproficient.

The nation's largest metropolitan areas — including New York City, Los Angeles, Chicago and Dallas — would all stand to gain at, or just above, 10% of their GDP by eradicating illiteracy. In these cases, nonproficiency rates are similar to the national average, and local gains from literacy are expected to be higher, given the lucrative job opportunities for people with high levels of literacy. Among the top 10 largest metro areas, only Boston would gain less than 10% (7%), largely because it has relatively low levels of nonproficiency (44%).

Even with the variation noted, the gains from eradicating illiteracy would be large and noteworthy in every corner of the United States.

Frequently Asked Questions

1 **What do we mean by illiteracy?**

This report defines illiteracy as a lack of proficiency on the PIAAC, an internationally validated literacy exam. Adults who score below Level 3 for literacy are not considered proficient and are defined as at least partially illiterate in this study. Adults below or at Level-1 literacy may struggle to understand texts beyond filling out basic forms. Drawing inferences or combining multiple sources of texts is likely too difficult. Adults at Level 2 can read well enough to evaluate product reviews and perform other tasks that require comparisons and simple inferences, but they are unlikely to correctly evaluate the reliability of texts or draw sophisticated inferences. Adults at Level 3 and above are considered fully literate in this study. They can reliably evaluate sources, as well as infer sophisticated meaning and complex ideas from written sources.

2 **How does this analysis account for differences in labor markets and the cost of living between U.S. counties, metropolitan areas and states?**

Housing costs and other expenses differ widely across U.S. counties, as do incomes. In the urban economics literature, which specializes in explaining the economics of cities and geographic differences, income variances are largely attributed to productivity differences between workers, defined as the amount of revenue a worker can directly or indirectly contribute to by carrying out his or her job. Productivity, in this sense, depends, in part, on the wealth of local communities. In richer areas, for example, restaurants can charge high prices for luxury experiences, whereas in poor communities, there would not be enough customers to sustain such services. Thus, the median worker in New York City earns 2.4 times more than the typical worker in Marion County, Ohio.

When estimating gains from literacy, the best way to incorporate these differences would be to observe how income varies by literacy level in each county. Those data, however, are not available since the literacy estimates are indirectly estimated and not observed alongside individual income in a meaningfully large sample of adults in enough counties.

This approach, therefore, uses a different method. Using the national effect of income gains as a benchmark, it converts the national effect to the local labor market by multiplying the ratio of the bachelor's degree premium in the county to the national bachelor's degree premium. The national bachelor's degree premium is $22,204 in 2017 USD, and is calculated by subtracting the median earnings of high school graduates from the median earnings of bachelor's degree holders who have not gone to graduate school ($52,019-$29,815).

The earnings gap between bachelor's degree holders and high school graduates is a good proxy for the earnings gap between people at Level 3 and people at lower levels because roughly two-thirds of bachelor's degree holders are at or above Level 3 (with the largest proportion at Level 3), whereas two-thirds of high school graduates with no postsecondary education are at or below Level 2. The Census Bureau's 2017 5-Year American Community Survey was used for this analysis and has large enough sample sizes for nearly every county. The same analysis was conducted using metropolitan areas and states.

The equation is shown below:

$$\text{Local adjustment factor} = \frac{\textit{Median income of bachelor's degree holders minus income of high school graduates in county}}{\textit{Median income of bachelor's degree holders minus income of high school graduates in USA}}$$

This adjustment factor is multiplied by the expected income gains for each county (or state or metropolitan statistical area). Thus, the gains are magnified somewhat in higher-income counties (like New York), relative to

low-income counties (like Marion County, Ohio), because highly educated workers in higher-income counties typically enjoy a larger earnings premium.

In 77 very small counties (with less than 15,000 adults, aged 25 and older), the above method yielded missing or negative estimates for the adjustment factor. An increase in literacy cannot cause a decrease in earnings potential, so an alternative method was developed. Instead of using the earnings gap between those with a bachelor's degree and those with a high school diploma, Gallup made the following substitutions: If the preferred gap (i.e., earnings for bachelor's degree holders, less earnings of high school graduates who have not attended college) was missing or negative, it was replaced with the gap between those with a bachelor's degree and those with less than a high school diploma. If this was still missing or negative, it was replaced it with the gap between those with a graduate degree and those with a high school diploma. Finally, if this was still missing or negative, Gallup multiplied the median income for all adults (which was not missing for any county) by 1.33, which is the average ratio of income for those with a bachelor's degree to those with a high school diploma.

Thus, the final estimate of income gains from moving all adults to proficiency in literacy for each county (c) is represented in the formula below, where P stands for population, and the subscripts represent the change in the level of literacy.

$$Aggregation\ Gains_c = (Gains_{1-3} \times P_{1-3,c} \times Local\ adjustment_c + Gains_{2-3} \times P_{2-3,c} \times Local\ adjustment_c)$$

Thus, variation in income gains are driven by variation in the population of adults with less than Level-3 literacy and the local adjustment factor.

At the county level, the local adjustment factor essentially eliminated any earnings gains from literacy in a handful of small counties, such as Bollinger County, Missouri, and Tillamook County, Oregon. One way to interpret this is that the local labor market does not have jobs that consistently reward higher levels of education. In reality, people with higher levels of literacy could potentially start their own successful business, so this is, in many ways, a conservative estimate, but it respects the fact that some rural parts of the country do not generate many economic opportunities for people with high levels of education or literacy.

At the metropolitan level, the Missoula, Montana metropolitan area has the lowest adjustment factor (.26), meaning that income gains from literacy are expected to be roughly one-quarter of national income gains. At the other end, San Jose's gains are 2.1 times the national average, indicating very high rewards for literacy.

At the state level, the variation is less extreme. Gains from literacy in Vermont (the state with the lowest adjustment factor) are only half the national average, and in Washington, D.C. — which has the highest adjustment factor — gains are expected to be about 1.6 times the national average.

3 How did the U.S. Department of Labor estimate literacy rates in counties where few or no people took the PIAAC exam?

The U.S. Department of Education combined assessment data from three sample waves (2012, 2014 and 2017), using data from 12,330 respondents living in 185 counties. The research team then modelled the literacy scores (and numeracy scores, which are not a focus of this report), which means they gathered a large amount of data about each respondent and his or her county to predict that respondent's literacy score.[4]

In choosing which variables to consider as predictors of county-level literacy, the research team used 70 county-level variables and 20 state-level variables (those available and derived from the research literature). The goal was to prune the number by seeing which variables best fit the data but were not highly correlated with another variable in the model. The final list of variables included the share of county population with less than a high school education, the share with a postsecondary education, the poverty rate, the share of the population that is Black, the share of population that is Hispanic, the health insurance rate, and the percentage of workers in nonprofessional service occupations.

As the U.S. Department of Education research team says, the PIAAC county and state estimates can be described as "predictions of how the adults in a state or county would have performed had they been administered the PIAAC assessment." We can be confident in these predictions because the characteristics of people living in each place are highly predictive of actual scores, and the predicted results were not significantly different than the actual results for the counties where direct estimates were available. Nonetheless, because of variation in the quality of learning opportunities in these areas and other subtle differences in family or population characteristics, we would expect that some counties would perform somewhat better or worse on the PIAAC exam if a representative sample of adults actually took it.

4   How does literacy relate to other factors that affect income?

It is well known in the labor economics literature that education, age, gender, race and other demographic characteristics strongly predict income. Since literacy is very highly correlated with educational attainment — both because more literate people pursue higher levels of education and because educational attainment increases literacy — it is inappropriate to adjust earnings gains from literacy by education. Doing so would understate the gains from literacy.

Alternatively, if a young Black woman living in a central city increases her literacy score, it is unlikely that her labor market gains will be the same as an older White man living in the suburbs of a large city. Enhancing literacy does not change someone's birthday, gender or race, and it will not necessarily change the city they live in. It will certainly not change that person's early-life experiences and upbringing. PIAAC collects data on these characteristics, including how many books respondents had in their childhood home and the educational attainment of their parents. These things all predict differences in income, regardless of literacy, so it is appropriate to hold them constant in performing the simulation exercise of increasing literacy. Controlling for these factors, however, only explains some of the gains from literacy, as can be seen by comparing the first and second rows in Table 1.

Table 1 also compares the raw differences in income by literacy group to the preferred demographically adjusted estimates, as well as the differences when adjusting for demographics and educational attainment. The last row is presented only as a comparison, however, because an individual's educational attainment would likely increase if their literacy increased. The point is that much of the gains of literacy come through education, but education is an incomplete proxy for literacy. Even people with identical degrees earn substantially more if they have higher levels of literacy proficiency. The middle row (the demographically adjusted estimates) serves as the basis for the modelling work described above and form the baseline estimates for the income gains of each group.

Table 1. Estimated Income Differences Between U.S. Adults With Difference Levels of Literacy, Before and After Adjusting for Other Observable Characteristics

	Level 0 or 1 to Level 3	Level 2 to Level 3
Raw mean difference	$28,870	$15,400
Differences adjusting for demographics	$23,979	$13,193
Differences adjusting for demographics and educational attainment	$14,708	$7,786

Demographic variables include gender, nativity, a cubic in age, binary variables for cities, suburbs and towns (relative to rural areas), controls for race and Hispanic ethnicity, binary variables for whether the respondents parents went to college or did not have a high school diploma, and binary variables for the number of books in the respondents childhood home. The controls for educational attainment also control for health status, which was not significant. The data are from the 2017 PIAAC with reported monthly earnings converted into annual earnings in 2020 USD.

5   What is the relationship between income and GDP?

This simulation compares hypothetical income gains to GDP, which is conceptually valid because gains in income directly increase the GDP. The Bureau of Economic Analysis produces the U.S. Gross Domestic Product estimates. BEA defines GDP as "the market value of the goods, services and structures produced by the economy in a given period."[5] There are several ways of measuring the GDP that are conceptually the same but require different methods and data sources. One way of measuring the GDP for a year is to measure all income earned through work (i.e., earned during production) that year. This adds the wages and salaries of workers to business profits and taxes (and the depreciation of equipment and structures).[6] Thus, income earned through work — including business profits — can be thought of as a component of GDP.

Notes

---

[1] Card, David. "The causal effect of education on earnings." In *Handbook of labor economics*, vol. 3, pp. 1801-1863. Elsevier, 1999. "U.S. Department of Education, Office of Career, Technical, and Adult Education. (2015, February). Making Skills Everyone's Business: A Call to Transform Adult Learning in the United States."

[2] Mankiw, N. Gregory, David Romer, and David N. Weil. "A contribution to the empirics of economic growth." *The quarterly journal of economics* 107, no. 2 (1992): 407-437; Jones, Charles I. "The facts of economic growth." In *Handbook of macroeconomics*, vol. 2, pp. 3-69. Elsevier, 2016; Hanushek, Eric A., and Ludger Woessmann. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of economic growth* 17, no. 4 (2012): 267-321.

[3] These data are available here, **https://nces.ed.gov/surveys/piaac/state-county-estimates.asp#4**

[4] Krenzke, T., Mohadjer, L., Li, J., Erciulescu, A., Fay, R., Ren, W., Van de Kerckhove, W., Li, L., and Rao, J.N.K. (2020). Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report (NCES 2020-225). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020225.

[5] U.S. Bureau of Economic Analysis, National Income and Product Accounts Handbook, Chapter 2: Fundamental Concepts, **https://www.bea.gov/system/files/2019-12/Chapter-2.pdf** (accessed May 11, 2020)

[6] The consumption of fixed capital is the most conceptually challenging concept in gross domestic income accounting. An example may help. If a truck-driver puts $5,000 worth of wear-and-tear on his or her truck over the course of a year, this is added to the value of GDP because it is akin to the wages paid to the machine that are not otherwise accounted for, whereas the cost of buying the truck is counted as GDP in the form of income accruing to the previous owner, and the costs of repairing the truck are counted as GDP when the repair service technician is paid.