The Occurrence of Low Literacy Among Adults in U.S. States and Counties

Research and Development Report

2010

Sheida White
National Center for Education Statistics

Tom Krenzke
Westat

Dan Sherman
American Institutes for Research

Sheida White
Project Officer
National Center for Education Statistics

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Statistics
Stuart Kerachsky
Deputy Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

> National Center for Education Statistics
> Institute of Education Sciences
> U.S. Department of Education
> 1990 K Street NW
> Washington, DC 20006-5651

Month 2010

The NCES World Wide Web Home Page address is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog address is http://nces.ed.gov/pubsearch.

Suggested Citation
White, S., Krenzke, T., and Sherman, D. (2010). *The Occurrence of Low Literacy Among Adults in U.S. States and Counties* (NCES 2010). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

For ordering information on this report, write to
> U.S. Department of Education
> ED Pubs
> P.O. Box 22207
> Alexandria, VA 22304

or call toll free 1-877-4ED-Pubs or order online at http://www.edpubs.gov.

> **Content Contact**
> Sheida White
> (202) 502-7473
> sheida.white@ed.gov

Executive Summary

In 2009, the National Center for Education Statistics (NCES) published a technical report titled *Indirect County and State Estimates of the Percentage of Adults at the Lowest Literacy Level for 1992 and 2003*. NCES also published a corresponding online tool (http://nces.ed.gov/naal/estimates/index.aspx) that allows users to compare estimates of the percentage of adults with the lowest level of prose literacy for any two states or counties—or to compare the estimates for 1992 and 2003 by jurisdiction.

This report conveys in nontechnical terms the statistical methodology used to develop the estimates. It also provides major findings for all 50 states and the District of Columbia, a profile of adults lacking *Basic* prose literacy, and a description of various potential users and usages of the findings. By way of illustrating how literacy estimates from the report and tool may be interpreted and used, examples from three jurisdictions—the District of Columbia, California, and Connecticut—are provided.

While the statistical procedure used for state and county estimates was found to be a valid approach to estimation, the margins of error associated with county-level estimates are quite broad compared to state-level estimates. Policymakers must keep this in mind as they use the online tool on the NCES website.

This set of estimates is expected to provide information about the literacy of adults, ages 16 and older, lacking *Basic* prose literacy in English (i.e., those who have *Below Basic* prose literacy and those who are unable to participate in the assessment due to language barriers). These individuals may be able to read simple words and phrases, but are generally unable to read and comprehend connected text in English, such as a newspaper story.

Selected findings include the following:

- The estimates of the percentage of adults lacking *Basic* prose literacy across states in 2003 range from 6 percent to 23 percent.

- The national direct estimate in 2003 of the percentage of adults lacking *Basic* prose literacy in English was 14.5 percent—or about 32 million (1 in 7) adults.

- The corresponding figure from the 1992 National Adult Literacy Survey is very similar: 14.7 percent—or about 29 million[1] (1 in 7) adults, so there is no

---

[1] This number was based on the sum of the population estimates of adults ages 16 and older across all states in 1992 (196 million) obtained from the NAAL web tool multiplied by 0.147. The same approach was applied to arrive at the 2003 estimate of 32 million.

statistically significant change during the period 1992 to 2003 in the proportion having low literacy.

- While there was no significant change for the nation during the period 1992 to 2003 in adults' low literacy, there were significant changes for a few states. Three states (Kentucky, Missouri, and Rhode Island) had a significant increase of literacy rates during the past decade and two states (California and New York) had a significant decrease of literacy rates.

- Overall, 10 percent of apparent differences between survey years at state level were statistically significant—a higher rate than that detectable at the county level (1 percent).

- About 5 percent of U.S. adults (representing approximately 11 million adults) were estimated to be *Nonliterate in English*, which encompassed two groups:

  1. One *Nonliterate in English* group (representing about 2 percent of U.S. adults, or 4 million adults) knew neither English nor Spanish (the other language spoken by interviewers in most areas) and therefore was unable to participate in the assessment at all.

  2. The other group included in the *Nonliterate in English* category (representing 3 percent of U.S. adults, or 7 million adults) performed poorly on a set of basic screening tasks, indicating that they would be able to perform few, if any, tasks in the main National Assessment of Adult Literacy.

- About 63 percent of adults *Nonliterate in English* were Hispanic (compared to 12 percent in the nation), and 69 percent of these were Mexican (compared to 58 percent in the nation) based on Kutner et al. (2006, table D 2-9).

- More than half (55 percent) of those in the *Below Basic* prose literacy group (compared to 15 percent of all adults) did not complete high school.

Acknowledgments

This page is intentionally blank.

Contents

List of Exhibits

List of Figures

List of Tables

Introduction

The Need for Information to Inform Planning for Literacy Services

According to the U.S. Department of Education's Office of Vocational and Adult Education 2007 report, *Adult Education Annual Report to Congress Year 2004-05,* approximately 2.6 million adults in the country were enrolled in state-funded adult education programs in the program year 2004-05. Of these, 39 percent were enrolled in adult basic education programs, 16 percent in adult secondary education programs, and 44 percent in English literacy programs. These numbers do not take into account adults served by other social service agencies, community colleges, and volunteer-operated literacy programs; nor does it take into account illegal immigrants, who were not represented in that sample.

> *Current data suggest that the need for adult education in general—and basic literacy education for adults in particular—is not being met.*

More current data suggest that the need for adult education in general—and basic literacy education for adults in particular—is not being met. A report, *Adult Student Waiting List*, published in 2006 by the National Council of State Directors of Adult Education, revealed that the great majority of programs have waiting lists. Although the extent to which demand outstrips capacity varies by state, in many states adults must wait months to access services. In Rhode Island, for example, the majority of program participants waited a year or longer.

States distribute state and federal funds competitively to local adult education providers, a network that includes a variety of local agencies—school districts, community colleges, and community-based and volunteer literacy organizations. Many adult education programs also work with state and local welfare agencies to provide instruction to adults receiving Temporary Assistance for Needy Families benefits who need to achieve *Basic* prose literacy. In short, there are many claims on funds that support adult education programs.

Providing policymakers with state and county estimates of percentages of adults lacking *Basic* prose literacy is the first step in helping them to better assess the educational needs of their respective states and counties. Such estimates provide the context of need and can indicate to some extent whether state and local adult educational programs and policies are having the desired outcomes, allowing officials to make the case for new initiatives or policy changes.

Indirect Estimates of Adult Literacy

In early 2009, the U.S. Department of Education's National Center for Education Statistics (NCES) published a report titled *Indirect County and State Estimates of the Percentage of Adults at the Lowest Literacy Level for 1992 and 2003*, which gave estimates for those 2 years of the percentage of adults in each state and county who lack *Basic* prose literacy in English. The report and an accompanying website (http://nces.ed.gov/naal/estimates/index.aspx) present estimates for all 50 states, the

3,141 counties within those states, and the District of Columbia. The estimates are derived from a statistical analysis of the 2003 National Assessment of Adult Literacy (NAAL) and the 1992 National Adult Literacy Survey (NALS).

The 2003 NAAL assessed a nationally representative sample of adults age 16 and older. While sufficient for estimating levels of adult literacy for the nation as a whole, the sample size was not large enough to provide direct estimates of literacy for most individual states and counties. Yet policymakers and educators throughout the country have a need for such information for their jurisdictions, particularly regarding that portion of the adult population with the lowest skill levels.

> *Providing policymakers with state and county estimates of percentages of adults lacking Basic prose literacy is the first step in helping them to better assess the educational needs of their respective states and counties.*

NCES responded to this need by applying a sophisticated statistical estimation model that could provide estimates of the percentage of adults with low-level literacy skills for all states and their counties. The model used information on both (1) the actual percentages of such adults in those states and counties that had NAAL sample cases, and (2) demographic characteristics such as low educational attainment, foreign-born status, poverty, certain Census geographic divisions, and race/ethnicity—called *predictor variables*—for all counties from the 2000 Census of the Population. These predictor variables, as a group, were correlated with the percentage of adults lacking *Basic* prose literacy in English. Once refined and evaluated, the 2003 model was then applied to the 1992 NALS using the 1990 Census of the Population demographic data to obtain similar predictor variables for the 1992 model.[2]

Because these estimates of the percentage of adults with low-level literacy skills are made using statistical models rather than actual counts, they are called *indirect estimates*. And because they are estimates pertaining to states and counties, rather than national estimates, they are also referred to as *small area estimates*. Finally, because the relationships between literacy and predictor variables are described in the form of mathematical models, the estimates are considered *model-based*.

The models were used to predict small area estimates of the percentage of adults who lack *Basic* prose literacy in English for all states and counties for both assessment years. This report displays the state estimates; the website has the county estimates, plus a web

---

[2] The decennial censuses were used rather than demographic surveys closer in time to the literacy assessments because their data were more robust. The report lists all surveys and variables that were considered and tested.

tool for users to make state and county comparisons within and between years. In the absence of any other literacy assessment data or a direct sample, these indirect estimates present the best available picture of the percentage of adults lacking *Basic* prose literacy in every state and county. They are predictions of the percentage of such adults in any given state or county that would have been directly estimated had a large enough sample been administered in the prose literacy assessment.

Purpose of This Report

While the full technical report (Mohadjer et al. 2009) describes in detail how the estimates for these jurisdictions were derived, this report seeks to convey the findings and other essentials about this study in nontechnical language and in an easily accessible format to reach a broad audience of leaders responsible for adult basic education in the public and private sectors.

Until now, those who make policy decisions on adult literacy programs below the national level have lacked a methodological approach that both provides (1) an estimate of the prevalence of very low literacy and (2) information about the estimate's level of precision.[3] Nor have they been able to make comparisons of literacy competencies over time. This report describes the development and results of a new online tool that offers educators, state directors of adult education, and other decisionmakers estimates of levels of low literacy in their locales, and the need for adult basic literacy planning among their constituencies.

> *The statistical estimation model used information on both (1) the actual percentages of adults lacking Basic prose literacy in those states and counties that had NAAL sample cases, and (2) demographic characteristics—such as low educational attainment, foreign-born status, and poverty—called predictor variables.*

---

[3] Prior to development of the method introduced in this report, Reder (1997) created a statistical model using the 1992 NALS dataset to generate literacy estimates for states and counties (and other geographic areas). The present NCES model reflects extensive research that has taken place in the interim in the field of small area estimation to capture more factors that can affect estimates, such as sampling error and model error, thereby providing more reliable estimates of the margin of error in the estimates. That is, the precision measure now gives a better picture of the uncertainty associated with such model-based estimates. As a result, the width of the credible intervals (explained under the section Credible Intervals later in this report) is larger (median of 14.5 percentage points for counties as reported in Mohadjer et al. (2009) as compared to a median of 6 percentage points as reported in an archive report on Reder's work on the Comprehensive Adult Student Assessment Systems (CASAS) website. More information can be found at https://www.casas.org/home/index.cfm?fuseaction=home.showContent& MapID=2800.

Although imprecise—especially for counties—in the absence of any other literacy assessment data, these indirect estimates present the best available picture of the percentage of adults lacking Basic prose literacy.

In addition to conveying the technical report's findings and methodology in nontechnical language, the current publication provides information not available in the technical report, including

- the need for indirect estimates;

- ways to interpret the findings accompanied by state examples;

- a profile of adults lacking *Basic* prose literacy;

- users and usages of the findings; and

- limitations and future directions.

Methodology

We begin our discussion of methodology with a review of what was learned from NAAL at the national level about adults ages 16 and older who lack the ability to successfully complete basic everyday English prose literacy tasks, such as comprehending a news story and using that information to accomplish daily goals.

Profile of Adults Lacking *Basic* Prose Literacy

As described in previous NAAL reports (e.g., Kutner et al. 2006, White and Dillow 2006), adults lacking *Basic* prose literacy in English are arranged in two categories: (1) those who are *Nonliterate in English*, and (2) those who have *Below Basic* prose literacy.

Nonliterate in English

As shown in figure 1, about 5 percent of U.S. adults (representing approximately 11 million adults) were not literate in English, although they may have been literate in some other language. This category encompasses (1) respondents (representing about 2 percent of U.S. adults, or 4 million adults) who knew neither English nor Spanish (the other language spoken by interviewers in most areas) and therefore were unable to participate in the assessment at all; and (2) respondents (representing 3 percent of U.S. adults, or 7 million adults) who participated in the assessment but scored very low on the simple core questions placing them at the bottom of the *Below Basic* category in prose literacy (i.e., in the NAAL supplemental assessment). This group represents adults who have a great deal of difficulty reading in English. About 63 percent of adults *Nonliterate in English* were

Figure 1.   Percentage of adults in *Below Basic* prose literacy and *Nonliterate in English* levels: 2003

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

Hispanic (compared with 12 percent in the nation) and 69 percent of those were Mexican (compared with 58 percent in the nation). No demographics are available for the 2 percent of adults in 2003 that were unable to participate at all because of language barriers.

Below Basic prose literacy

While those in the *Nonliterate in English* group have a great deal of difficulty reading in English, the *Below Basic* prose literacy group can read a little but not very well. While those at the bottom of this category are able to identify letters, numbers, and simple words and phrases, most—including those at the upper end of the *Below Basic* level—are unable to read and comprehend connected text in English, such as a newspaper story. More specifically, they lack basic reading skills, such as the ability to decode unfamiliar words that are printed or written, to recognize familiar words that are printed or written, or to read with fluency (i.e., with speed and ease). Demographically speaking, more than half (55 percent) of these respondents had less than or some high school, compared to 15 percent of all adults who have received this level of schooling.

> *The estimates of the percentage of adults lacking Basic prose literacy across states run from 6 percent to 23 percent.*

Developing the Indirect Estimation Model

While the NAAL provided direct estimates of low literacy among adults at the national level, the main purpose of developing the indirect estimation model was to report on the percentage of adults lacking *Basic* prose literacy in English at the state and county levels. The model was developed and tested in the manner described below.

Data sources for the indirect estimates

The foundation of the indirect estimation model was the 2003 NAAL data. To design the indirect estimation model these data were used in conjunction with more than 100 variables drawn primarily from the United States Census 2000 county-level data.

Literacy data from sampled counties in the 2003 NAAL and 1992 NALS

The NAAL sample was designed for producing national level estimates with adequate precision. A design to produce direct state and county estimates would demand substantial increases to both burden and cost. Indeed, for most counties and even for most states, the NAAL sample size is not large enough to provide direct estimates of literacy with adequate precision, raising the need for small area estimates.

Participants in NAAL responded to a series of tasks that measured a person's ability to use and comprehend printed and written material represented on three scales of literacy: prose, document, and quantitative. Respondents could demonstrate performance at successive competency levels called: *Below Basic, Basic, Intermediate,* and *Proficient.* These assessments produced standard estimates, or *direct estimates*, of literacy for the

nation as a whole and for major subgroups that included groupings by age, education, race/ethnicity, and regions.

The task of developing a statistical model from NAAL for indirect estimates began with deciding upon (1) which assessment results to focus, and (2) which performance level to report. Comprehending and using prose text such as news stories is a primary facet of literacy as it enables adults to comprehend and use not only single words and phrases, but also sentences and paragraphs. The performance level to be estimated was decided as *Below Basic*[4] because this group is most in need of basic literacy education. This publication refers to this group as *adults lacking Basic prose literacy*.

It is relevant to the creation of the estimation model that within each of the two literacy assessments, the 2003 NAAL and the 1992 NALS, some states paid to have additional sampling so that literacy levels of their populations could be more precisely estimated. Six states in 2003—Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma—chose to have NCES conduct a State Assessment of Adult Literacy (SAAL) in their jurisdictions. In 1992, 11 states—California, Illinois, Indiana, Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, and Washington—had voluntarily participated in the State Adult Literacy Survey (SALS).

Thus for those years, these states have direct estimates of adult literacy for their populations. Later, these direct state estimates became integral to the evaluation of the indirect estimation model, because the data were used to validate the projections for the participating states and then to help confirm the validity of the model's projections for the other states as well.

NAAL sampled about 18,500 adults ages 16 and older residing in households in the 50 states and the District of Columbia in 2003. This included the six state-level direct samples mentioned above, which had an aggregate sample size of about 5,800 adults. For the nation, Black and Hispanic adults were oversampled to gain greater precision of estimates.

The NAAL sampling process selected individual adult respondents from households (not group homes or institutions) across 342 counties. Interviewers met with each respondent to determine his or her demographic characteristics, educational background, and other literacy-related factors. They then presented the respondents with a set of literacy tasks covering the three components already noted. These tasks simulated the kind of written prose, document, and quantitative materials that adults encounter on a regular basis. Each respondent received only a small sample of the tasks, however, so total literacy estimates

---

[4] There is a significant difference in basic reading skills (i.e., decoding and word recognition) between the literacy of adults scoring at the *Below Basic* and *Basic* levels. For example, when presented with health-related texts, adults at the *Below Basic* level read 66-125 words correctly per minute, whereas those at the *Basic* level read 139-150 words correctly per minute (White 2008).

were arrived at using a statistical procedure for combining responses from the entire sample to project a general outcome. Of the sample, 2 percent could not be assessed because they could not communicate in either English or Spanish. These indirect estimates include them in the group that lacks *Basic* prose literacy in English. Neither the 2003 NAAL nor 1992 NALS included in their numbers those who could not take the assessment due to a language barrier, whereas the indirect estimation model described here does include them. Thus the model's low English prose literacy group is somewhat larger than NAAL's *Below Basic* literacy category, so the two are not comparable.

Direct estimates of the percentage of this group were obtained for 264 of the 342 counties sampled in the 2003 NAAL, or 8 percent of the U.S. counties. (The other 78 counties were excluded because of technical deficiencies, such as having fewer than five participating adults.) Due to the focus on creating national-level estimates and the cost of interviewing in every county, most counties as well as 12 states do not have a NAAL sample, and therefore the resulting indirect estimates for these counties and states are purely dependent on the model. In the 1992 NALS, direct estimates of low prose literacy were based on 368 sampled counties, or 12 percent of the U.S. counties. The median county sample size was 35 adults in the 2003 NAAL and 41 adults in the 1992 NALS. The percentage lacking *Basic* prose literacy in English was estimated for each of these counties using the prose items in the assessment and with scores below 210.

The model used the Census predictor variables, together with direct county estimates, to predict low adult literacy estimates for these nonsample counties, taking into account the demographics of these counties, as will be explained further, below. For counties with large samples, the estimates were influenced more by the direct estimates, and for counties with small samples, the estimates were influenced much more by the model. Figure 2 displays a scatter chart that plots the residual in relationship to the county

sample sizes, for the 264 counties with direct estimates. The residual is the difference between the predictions and the direct county estimates. Therefore, the residual can be either positive (larger prediction than direct estimate) or negative (smaller prediction than direct estimate). The chart shows that the larger the sample, the smaller the residual, which means that the estimates depend more on the direct estimates for larger samples than on the model.

The predictions for the remaining *nonsample* counties are based solely on the model, using the associations that had been established between the direct estimates and the predictor variables among the sample counties, and applying those associations to the nonsample counties and their predictor variables.

Demographic variables most correlated with literacy

As mentioned earlier, formulating indirect estimations for both sampled and nonsampled counties and states requires demographic variables that are effective predictors of low English prose literacy and that are measured consistently across all states and counties. To design the indirect estimation model, more than 100 variables drawn primarily from the Census 2000 county-level data were analyzed. These demographic variables were assumed to have predictor values for low literacy. They included level of education,

English-speaking capability, immigration, racial and ethnic minority status, age, employment status, type of employment, urban/rural status, and poverty status.

Once the direct county estimates from the 2003 NAAL were computed and once the potential predictor variables of low literacy were selected, they were tested to see if there was a statistical relationship between them and literacy (e.g., between education attainment and literacy). These resulting correlations were critical in establishing the indirect model. For example, if the NAAL data showed an association between percentage of adults with low literacy and low education attainment across the 264 sampled counties, as indeed it did, the percentage of adults with such low literacy levels in nonsample counties could be projected based, in part, on their education attainment. The list was reduced after conducting analyses that identified the strongest associations between the demographic variables and the likelihood of low literacy rates.

The following demographic variables emerged, as a group, in the final set of estimate variables and 2003 indirect estimation model. They were the percentage of the population that

- was foreign born and stayed in the United States for 0-20 years;
- had a high school education or less;
- was Black or Hispanic; and
- was below the 150 percent poverty line.

In addition, two Census indicators identifying location were created and incorporated into the model because of the differences in proficiency found in these areas during the model building process. One indicator variable identified was those living in New England or in the North Central United States. This indicator was added to account for some unexplained differences from effects other than country of birth, education, race, and poverty. To account for the sample design, which includes an increased sample size in SAAL states (i.e., states that participated in the State Assessment of Adult Literacy), a second indicator variable was introduced to identify counties that were associated with SAAL states. Once these variables were considered, it was found that additional variables added little to the predictive power of the model.

The final list—which together accounted for close to 40 percent of the variation in the percentage lacking *Basic* prose literacy—was then evaluated using relevant statistical diagnostic tests used in the field of small area estimation.[5]
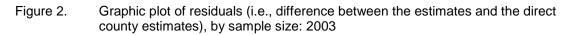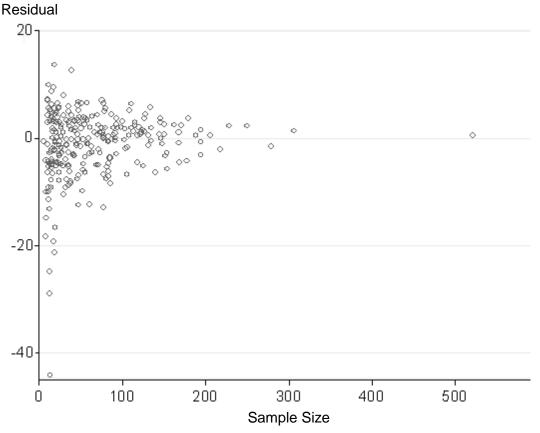
---

[5] Details about the model evaluation process can be found in chapter 5 of Mohadjer et al. (2009). See also, Gelman et al. (2004).

Evaluating the Indirect Estimation Model

Before the model could be implemented, different tests of "fit" were conducted. Alternative models, nine in total, were created, with different potential predictor variables added to the core predictors (percent foreign-born in the country for 0-20 years, percent with a high school education or less, percent Black or Hispanic, and a census division indicator) that were settled upon early in the process. These were added either because other research suggested their value as being associated with low literacy, or because it was hypothesized that they could strengthen the model. It was this evaluation process that resulted in the addition of the percentage in poverty to the final NAAL model.

Another evaluation approach was aimed at determining if the model performed as expected. That is, for counties with large samples, the estimates should be influenced more by the direct estimates, and for counties with small samples, the estimates should be influenced much more by the model. Figure 2 displays a scatter chart that plots the residual in relationship to the county sample sizes, for the 264 counties with direct estimates. The residual is the difference between the estimates and the direct county estimates. Therefore, the residual can be either positive (larger estimate than direct estimate) or negative (smaller estimate than direct estimate). The chart shows that the larger the sample, the smaller the residual, which means that the estimates depend more on the direct estimates for larger samples than on the model.

Figure 2.   Graphic plot of residuals (i.e., difference between the estimates and the direct county estimates), by sample size: 2003

Residual

Sample Size

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

As another evaluation approach, the 2003 NAAL direct estimates for the six states, which agreed to bear the cost of participating in the SAAL, were also compared with aggregates of indirect county estimates generated by the model. As shown in table 1, which is extracted from table 5-3 in Mohadjer et al. (2009), there is no significant difference between the indirect and direct estimates for the six SAAL states. As another example, the indirect county estimates were aggregated to the nation, and the direct and indirect estimates of the percentage of adults lacking *Basic* prose literacy in English for the nation only differs by 0.09 percentage points. These checks help to confirm the validity of the model.

Table 1.   Comparison of aggregated indirect county estimates and direct estimates for percentage of adults lacking Basic prose literacy skills, by state: 2003

Subgroup (Source: Year)	Indirect estimate		Direct estimate			Percent-age point difference	Relative difference (percent)[1]
	Number of counties	Weighted estimate	Sample size	Estimate	Standard error		
State Assessment of Adult Literacy state (NAAL: 2003)							
Kentucky	120	12.2	1,500	11.4	1.00	0.7	6.1
Maryland	20	11.2	1,000	9.4	1.37	1.8	19.5
Massachusetts	10	9.9	1,000	10.7	1.43	-0.8	-7.2
Missouri	120	7.5	1,000	7.1	1.03	0.3	4.5
New York	60	22.1	1,700	20.6	1.86	1.5	7.1
Oklahoma	80	12.3	1,300	12.5	1.62	-0.3	-2.2

[1] The relative difference is computed as the difference divided by the direct estimate. Differences when conducting the relative difference using numbers shown in the table are due to rounding. The calculations were done on unrounded numbers.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

Adapting the Estimation Model to 1992 NALS Data

Subsequent steps in developing the model for the 1992 NALS were the same as for the 2003 model, namely direct estimates from the 368 sampled counties were computed, predictor variables were selected using the 1990 Census of the Population, and the model was evaluated. Thus state and county estimates of literacy for two time periods could be compared.

The same predictor variables used in the 2003 model were considered in fitting the 1992 model using the 1990 values for the variables that came from the 1990 Census. These variables were tested for their significant relationship with the rate of low prose literacy in 1992 NALS. This analysis results in a different set of predictor variables for 1992 than the set of predictor variables for 2003. The variables that emerged as final predictors were the percentage of the population that

There is no significant difference between the indirect and direct estimates for the six participating states.

- was nonnative speakers of English;
- had a high school education or less; and
- was Black or Hispanic.

A summary of the development of the model can be found in exhibit 1.

Exhibit 1.   Methodology at a glance: Development of the indirect estimation models

Literacy measure
- lacking *Basic* prose literacy in English (i.e., being unable to comprehend connected text in English, such as a newspaper story).

Data sources for the indirect estimates
- literacy data from sampled counties (264 in the 2003 NAAL); and
- demographic variables most correlated with literacy (e.g., race/ethnicity) based on the 2000 and 1990 Census of Population.

Method for developing the indirect estimation model for the 2003 NAAL
- correlated the direct estimates to the potential predictor variables of low literacy to identify the strongest relationships; the list of predictor variables were reduced from more than 100 to just a few:
  - percentage foreign-born and stayed in the United States 0-20 years;
  - percentage with high school education or less;
  - percentage Black or Hispanic;
  - percentage below the 150 percent poverty line;
  - New England and North Central census divisions; and
  - Indicator of whether state was included in the SAAL.

Model evaluation for the 2003 NAAL
- evaluated nine alternative models; this resulted in another predictor variable: poverty; and
- compared the 2003 NAAL direct estimates for the six participating states that paid to augment their samples, with aggregates of indirect county estimates generated by the model; the estimates were close, differing by less than 1 percentage point.

Method for developing the indirect estimation model for the 1992 NALS
- used the final variables in the 2003 model as a base and used stepwise regression to correlate direct estimates of literacy for 368 sampled NALS counties to these variables;
- found that breaking "Percentage of the population that was Black or Hispanic" into components improved correlation;
- tested alternative indicators of census divisions and found "New England" and "North Central" census divisions had highest correlation with direct estimates;
- model was based using the following variables that had statistically significant correlation with direct estimates of literacy:
  - percentage with high school education or less;
  - percentage Black;
  - percentage Hispanic;
  - percentage non-English speaking;
  - New England census and North Central division; and
  - indicator of whether state was included in the SALS.
- The 1992 estimates were internally evaluated against direct county estimates to assess their difference.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

Both models included predictive variables relating to educational attainment and race/ethnicity. It is unlikely that the discrepancy in the number of predictor variables significantly affected the trend, or even the precision of the 1992 indirect estimates—the median of the 95 percent credible interval width (see explanation of this term under Credible Intervals below) for all states is 7 percent in 1992 and 6 percent in 2003.

State Estimates

Once the models were finalized, they were employed to estimate the percentage of adults lacking *Basic* prose literacy in English for sampled and nonsampled counties. State estimates were derived from the indirect county estimates by combining the estimates for all counties in a given state, weighted by their population size for ages 16 and older. The state estimates incorporated the impact of the 2003 NAAL and 1992 NALS samples within the state, if they were available, as the county estimates were combined. In general, the smaller the number of counties with samples available from NAAL or NALS within a given state, the more dependent that state estimate was on the predictor variables. Thus, if the number of counties with a NAAL sample was small, and if the association was weak between the explanatory factors and the literacy estimate, then the less reliable the state-level estimates became.[6]

Credible Intervals

Each indirect estimate has an associated *credible interval* (or prediction interval), which includes the lower and upper boundaries that defines an interval where there is a .95 probability that it contains the true percentage of adults lacking *Basic* prose literacy in English. A credible interval captures more than the sampling error, which is captured in a traditional confidence interval. It also accounts for the model prediction error. Say for example that the model predicts that an area has 12 percent of its adults at the lowest literacy level, with a 95 percent credible interval of 5 percent to 25 percent. This means there is a 95 percent chance that the actual value, while not necessarily 12 percent, is somewhere between 5 percent and 25 percent. The smaller the range of the credible intervals, the more reliable the estimates are. Note that the predicted percentage of adults with low literacy is not necessarily the midpoint of the credible interval.

The credible intervals for the differences between any two *county* indirect estimates are relatively large (the median width in 2003 is 22), meaning there is imprecision in the estimates. In fact, when comparing across years for a single county, only 1 percent of the 1992 and 2003 county levels had significant differences. The imprecision of county estimates, in itself, contributes to the imprecision in the estimates of change between counties, or across years for a single county.

The indirect estimates for states are more precise than those for counties. The median of the 95 percent credible interval width for all states is 7 percent in 1992 and 6 percent in 2003, enabling better detection of statistically significant change between 1992 and 2003 for individual states. A user who is interested in a county that happens to have a wide

---

[6] For example, table 4-6 of Mohadjer et al. (2009) shows that the median credible interval width among states without a NAAL sample is 6.0, while the median credible interval widths ranged from 3.2 to 5.3 among the 6 SAAL states.

A user who is interested in a county that happens to have wide credible interval may consider other similar counties with similar characteristics or their state estimate to judge their literacy estimate.

credible interval may consider other similar counties with similar characteristics or their state estimate to judge their literacy estimate.

It must be clear at this point that, because the estimates have been generated by a model and not through a direct survey with a large sample from each small area, many of the estimates are imprecise. This does not mean they are wrong; it does mean the precision-level of such local area estimates is now known and that the estimates must be used with caution taking into account the credible intervals shown, especially when comparing one estimate to another.

This page is intentionally blank.

Findings

The estimates of the percentage of adults lacking *Basic* prose literacy across states run from 6 percent to 23 percent, as shown in table 2. The national direct estimate in 2003 of the percentage of adults lacking *Basic* prose literacy in English was 14.5 percent with a margin of error of 1.2 percentage points, indicating that the national estimate is very reliable. The 14.5 percent translates to about 32 million (or 1 in 7) adults lacking *Basic* prose literacy. The corresponding figure from the 1992 NALS is very similar: 14.7 percent, so there is no statistically significant change during the period 1992 to 2003 in the proportion having low literacy among adults 16 and older living in U.S. households.[7]

While there was no significant change for the nation during the period 1992 to 2003 in adults' low literacy, there were significant changes for a few states. Three states (Kentucky, Missouri, and Rhode Island) had a significant increase of literacy rates during the past decade, and two states (California and New York) had a significant decrease of literacy rates. Overall, 10 percent of states show statistically significant differences between survey years—a higher rate than that detectable for counties (1 percent).

In addition to state-level findings displayed in tables 2 and 3, a separate website (http://nces.ed.gov/naal/estimates/index.aspx) provides information for both the state and counties. The website, designed specifically for these indirect estimates, provides users with search and analytical capabilities to compare any two states or counties to each other or across years, and to determine if the difference is statistically significant.

By way of illustration, Minnesota had an estimated 9 percent of adults who lack *Basic* prose literacy in English in 1992 compared to an estimated 6 percent in 2003 (see tables 2 and 3). Whether the difference between the 2 years is statistically significant can be computed by the website user tool. After entering the state and dates, a table will be displayed showing the estimated difference and credible interval for the difference along with a statement of the results. If the difference is significant, the statement will note, "The difference between the estimated percentages of adults lacking *Basic* prose literacy skills (BPLS) is large enough to conclude, with at least a 95 percent chance, that there is a statistical difference between them." However, in the case of Minnesota, the following statement is displayed: "The difference between the estimated percentages of adults lacking BPLS is not large enough to conclude, with at least a 95 percent chance, that there is a statistical difference between the two years."

---

[7] This pool excludes residents of group dwellings such as nursing homes and armed forces personnel stationed elsewhere.

Table 2. Indirect estimates of the percent lacking *Basic* prose literacy skills and corresponding credible intervals, by state: 2003

States	Population size[1]	Percent lacking *Basic* prose literacy skills[2]	95 percent credible interval[3]	
			Lower bound	Upper bound
Alabama	3,400,000	15	11.8	19.4
Alaska	461,000	9	6.1	13.3
Arizona	4,080,000	13	9.6	18.1
Arkansas	2,040,000	14	10.2	17.2
California	26,030,000	23	20.3	26.2
Colorado	3,390,000	10	7.1	12.9
Connecticut	2,670,000	9	5.5	12.5
Delaware	619,000	11	6.6	16.4
District of Columbia	426,000	19	9.3	33.1
Florida	13,040,000	20	17	22.9
Georgia	6,366,000	17	14	20.7
Hawaii	944,000	16	11.5	22.2
Idaho	1,000,000	11	8	13.8
Illinois	9,510,000	13	10.4	16.6
Indiana	4,630,000	8	6.1	10.3
Iowa	2,250,000	7	5.3	10.1
Kansas	2,050,000	8	5.9	10.2
Kentucky	3,200,000	12	10.3	14.3
Louisiana	3,310,000	16	12.5	20.3
Maine	1,040,000	7	5.2	10.2
Maryland	4,190,000	11	9.1	13.7
Massachusetts	5,100,000	10	8.3	12.1
Michigan	7,630,000	8	6.2	11
Minnesota	3,900,000	6	4.1	8
Mississippi	2,120,000	16	11.9	20.8
Missouri	4,320,000	7	5.9	9.2
Montana	704,000	9	5.9	12.2
Nebraska	1,310,000	7	5.3	9.7
Nevada	1,670,000	16	9.5	25.3
New Hampshire	995,000	6	4	8.2
New Jersey	6,610,000	17	13.5	20.8
New Mexico	1,390,000	16	12.2	21.6
New York	15,060,000	22	19.7	25
North Carolina	6,280,000	14	11	16.5

See notes at the end of the table.

Table 2.   Indirect estimates of the percent lacking *Basic* prose literacy skills and corresponding credible intervals, by state: 2003—Continued

| States | Population size[1] | Percent lacking *Basic* prose literacy skills[2] | 95 percent credible interval[3] | |
			Lower bound	Upper bound
North Dakota	489,000	6	4.2	9
Ohio	8,720,000	9	7.2	12
Oklahoma	2,700,000	12	10.4	14.5
Oregon	2,710,000	10	7.3	13.9
Pennsylvania	9,560,000	13	10.2	15.5
Rhode Island	832,000	8	4.7	13.9
South Carolina	3,100,000	15	11.6	18.4
South Dakota	572,000	7	4.7	9.7
Tennessee	4,440,000	13	10.5	16.5
Texas	15,940,000	19	16.4	22.1
Utah	1,640,000	9	6.1	13.9
Vermont	485,000	7	4.4	9.4
Virginia	5,520,000	12	9.6	14.8
Washington	4,640,000	10	7.3	12.8
West Virginia	1,420,000	13	10.2	17.2
Wisconsin	4,190,000	7	5.1	9.9
Wyoming	382,000	9	6.2	12.2

[1] Estimated population size of persons ages 16 years and older in households in 2003.
[2] Those lacking *Basic* prose literacy skills include those who could not be tested due to language barriers and those who scored below the *Basic* level in prose.
[3] The estimated percent lacking *Basic* prose literacy skills is subject to uncertainty, as measured by the associated credible interval. The probability that the true value is contained between the lower and upper bound is .95.
NOTE: Population sizes are rounded.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

Table 3.   Indirect estimates of the percent lacking *Basic* prose literacy skills and
             corresponding credible intervals, by state: 1992

| States | Population size[1] | Percent lacking *Basic* prose literacy skills[2] | 95 percent credible interval[3] | |
			Lower bound	Upper bound
Alabama	3,190,000	21	14.5	27.4
Alaska	416,000	10	6	14
Arizona	2,950,000	13	9.4	17.5
Arkansas	1,840,000	19	12.9	25.8
California	23,230,000	15	11.8	17.9
Colorado	2,650,000	9	5.7	13
Connecticut	2,590,000	14	8.3	20.1
Delaware	536,000	12	7.8	15.8
District of Columbia	488,000	21	14.6	28.6
Florida	10,800,000	15	10.9	20.2
Georgia	5,110,000	18	12.8	24.8
Hawaii	889,000	18	13.9	23.2
Idaho	779,000	10	6.4	13.9
Illinois	8,930,000	15	12.3	18.2
Indiana	4,350,000	10	7.4	14
Iowa	2,160,000	7	4.3	9.9
Kansas	1,910,000	9	5.7	13
Kentucky	2,900,000	19	13.2	26.3
Louisiana	3,170,000	21	15.4	27.1
Maine	957,000	13	7.4	18.8
Maryland	3,790,000	12	8	17.2
Massachusetts	4,760,000	13	8.7	17.8
Michigan	7,200,000	12	8.5	16.2
Minnesota	3,390,000	9	5.4	12.1
Mississippi	1,950,000	25	17.9	34
Missouri	3,990,000	13	8.5	17.6
Montana	617,000	9	5.7	13.1
Nebraska	1,220,000	8	5.3	12.3
Nevada	1,040,000	13	9.7	17.7
New Hampshire	855,000	11	6.4	16.3
New Jersey	6,160,000	16	12.2	19.6
New Mexico	1,170,000	17	11.1	24.3
New York	14,190,000	16	12.9	20.1
North Carolina	5,380,000	18	12.6	24.6

See notes at the end of the table.

Table 3.   Indirect estimates of the percent lacking *Basic* prose literacy skills and
corresponding credible intervals, by state: 1992—Continued

States	Population size[1]	Percent lacking *Basic* prose literacy skills[2]	95 percent credible interval[3]	
			Lower bound	Upper bound
North Dakota	482,000	11	7.2	16.2
Ohio	8,450,000	12	8.5	15.4
Oklahoma	2,439,974	13	8.5	18.8
Oregon	2,300,000	10	6.2	13.3
Pennsylvania	9,440,000	13	9.8	17.3
Rhode Island	799,000	18	12.7	23.4
South Carolina	2,760,000	20	14	28.1
South Dakota	526,000	11	6.7	15.2
Tennessee	3,910,000	19	13	25.3
Texas	13,110,000	18	13.5	22.7
Utah	1,250,000	8	5.2	12.4
Vermont	439,000	11	6.4	15.9
Virginia	4,970,000	15	10.2	20.7
Washington	3,920,000	7	4.9	10
West Virginia	1,420,000	17	11.6	24.2
Wisconsin	3,820,000	10	6.1	13.6
Wyoming	342,000	9	5.3	12.4

[1] Estimated population size of persons ages 16 years and older in households in 1992.
[2] Those lacking *Basic* prose literacy skills include those who could not be tested due to
language barriers and those who scored below the *Basic* level in prose.
[3] The estimated percent lacking *Basic* prose literacy skills is subject to uncertainty, as measured
by the associated credible interval. The probability that the true value is contained between the
lower and upper bound is .95.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for
Education Statistics, 1992 National Adult Literacy Survey.

Inherent in the small area estimation model are predictor variables (poverty, education
attainment, race, foreign-born status) that help to capture the range in literacy skills as
measured by the percentage of adults with low English prose literacy. The association of
these predictor variables with direct estimates of low literacy is why they were selected
for the model from many other potential predictor variables. Consequently, this feature of
the model produces a range in low literacy estimates that is greater than among counties
than among states. For example, Webb County, Texas, has a higher than average
percentage of foreign-born adults (28 percent compared with the U.S. average of 12.5

percent); more adults with a high school education or less (60 percent vs. 45 percent U.S. average); a high Hispanic population (94.5 percent vs. 15 percent for the United States); and more poverty (25 percent of families live below poverty level compared with U.S. average of 9.6 percent).[8] The cumulative impact of each of these variables contributes to the high value for the percentage of adults who lack *Basic* prose literacy in English; in Webb County, the estimate is 48 percent, which is more than double the highest state-level estimate (23 percent in California).

The Census division indicator, which is part of the model, also contributes to the estimation of adult literacy in the county.[9] For example, living in the New England division or in one of the North Central Census divisions is associated with lower percentages of those with low *Basic* prose literacy—offsetting to some degree the upward direction of the projections caused by the predictor variables. This is reflected in the findings. For example, there is a big difference in survey estimates on low literacy between the group consisting of New England states (8.7 percent), East North Central states (9.6 percent), and West North Central states (7 percent) versus Middle Atlantic states (18.1 percent) and Pacific states (19.9 percent).

A website, http://nces.ed.gov/naal/ estimates /index.aspx, *provides information for both the state and counties.*

Individuals interested in getting the predictor variables for a county in their state (e.g., percentage poverty in 2000), for a state, or for the nation as a whole are encouraged to use the U.S. Census Bureau's web tool called the American FactFinder (http://factfinder.census.gov).

Three Examples: A District and Two States

Illustrations of how the model provides estimates for states and smaller jurisdictions are provided by taking a closer look at the District of Columbia, California, and Connecticut.

District of Columbia: An example of a smaller jurisdiction

As shown in table 4 below, the indirect estimate of the adult population ages 16 and older in the District of Columbia that lacks *Basic* prose literacy in 2003 is 19 percent. There is, however, a credible interval width of 24 percentage points associated with this estimate. This means there is a 95 percent chance that the true percentage of low literacy in California is between 20 percent and 26 percent. The indirect estimate for 1992 of the

---

[8] Figures from the U.S. Census Bureau's web tool, the American FactFinder (http://factfinder.census.gov).
[9] The Census divisions are (1) New England, (2) Middle Atlantic, (3) East North Central, (4) West North Central, (5) South Atlantic, (6) East South Central, (7) West South Central, (8) Mountain, and (9) Pacific.

percentage of the adult population ages 16 and older that lacks *Basic* prose literacy is 21 percent, with a credible interval width of 14 percentage points.

Table 4.　　Indirect estimate of adults in the District of Columbia lacking *Basic* prose literacy: 2003

| District | Population size | Percent lacking *Basic* prose literacy | 95 percent credible interval | |
			Lower bound	Upper bound
District of Columbia (1992)	488,000	21	14.6	28.6
District of Columbia (2003)	425,630	19	9.3	33.1

SOURCE: NAAL State & County Estimates of Low Literacy web tool at http://nces.ed.gov/naal/estimates/index.aspx.

To gain an understanding of how the model arrived at an estimate of 19 percent for 2003, those interested could use the same data sources used in the model. They would begin their exploration by using the U.S. Census Bureau's web tool, the American FactFinder (http://factfinder.census.gov) to search Census 2000 data for the percentage of adults having the demographic characteristics used as predictor variables in the 2003 statistical model. This search would reveal, for example, that the District of Columbia has a high percentage (67 percent) of Black and Hispanic adults (negatively correlated with literacy) when compared to the nation (24 percent). In combination with other Census 2000 variables used in the model (such as foreign-born), much of the variation from the national estimate would be explained. But, with a 24 percentage point credible interval, District of Columbia planners would need to use caution in their interpretation of the 2003 indirect estimate.

> *Individuals interested in getting the predictor variables for a state or a county in their state are encouraged to use the Census Bureau's web tool called the American FactFinder (*http://factfinder.census.gov*).*

California: An example of a large state

Consider California, a state with an estimated percentage of adults lacking *Basic* prose literacy that for 2003 is significantly higher than the nation as a whole. Indeed the California estimate for 2003 is substantially higher than its own estimate of the previous decade; it climbed 8 percentage points. Table 5, taken from the NAAL website (http://nces.ed.gov/naal/estimates/index.aspx), shows 1992 and 2003 estimates for California adults lacking *Basic* prose literacy. As in the District of Columbia example, the table also includes credible intervals to help gauge the precision of the estimate. In 2003, the estimated portion of the adult population lacking *Basic* prose literacy stands at 23 percent with a fairly small credible interval width of 6. This means there is a 95 percent chance that the true percentage of low literacy in California is between 20 percent and 26 percent. For 1992, the estimate is 15 percent, with about the same credible interval

width of 6. For that year, the true proportion would fall between 12 percent and 18 percent.

Table 5.   Indirect estimate of adults in California lacking *Basic* prose literacy: 1992 and 2003

| State | Population size | Percent lacking *Basic* prose literacy | 95 percent credible interval | |
			Lower bound	Upper bound
California (1992)	23,228,940	15	11.8	17.9
California (2003)	26,029,840	23	20.3	26.2

SOURCE: NAAL State & County Estimates of Low Literacy web tool at http://nces.ed.gov/naal/estimates/index.aspx.

Confident that the state numbers are reliable, literacy program planners might want to understand why the percentage of adults lacking *Basic* prose literacy in California is so much higher than the nation or why it has grown larger since 1992. The American FactFinder web tool (http://factfinder.census.gov) allows users to search for the percentage of adults having the demographic characteristics used as predictor variables because of their strong relationship with low literacy in the 2003 and 1992 statistical models. Table 6 reveals that, in California's case, the high percentage of Blacks and Hispanics (used in the model as a single variable) and of foreign-born residents is most at variance with the U.S. figures.

Table 6.   Comparison of predictor variables for California and the United States: Census 2000

Predictor variable	California	United States
Percent of adults ages 25+ with a high school education or less	43.3	48.2
Percent of Blacks/Hispanics	38.7	24.5
Percent of population below 150 percent poverty line	24.1	20.9
Percent of foreign-born people who stayed in the United States 0-20 years	18.2	7.7

SOURCE: U.S. Census Bureau, American FactFinder, Census 2000.

Table 7 shows a reduction in the percentage of people with a high school education or less in California during the decade from 1990 to 2000, which means Californians became better educated during that period. However, there was a concurrent increase in the state's percentage of Blacks and Hispanics—two groups commonly associated with lower literacy rates. This increase was larger than the increase in people going on to postsecondary education, so it may explain, at least in part, the rise in low literacy.

Table 7.   Comparison of predictor variables for California: 1990 and 2000

Demographics	1990	2000
Percent of adults ages 25+ with a high school education or less	46.1	43.3
Percent of Blacks/Hispanics	32.9	38.7

SOURCE: American FactFinder, Census 2000, 1990 Census of the Population.

Connecticut: An example of a small state

Connecticut is a state with an estimated percentage of adults lacking *Basic* prose literacy for 2003 that is lower than both the nation as a whole and its own estimate of the previous decade by about 5 percentage points. Table 8 (taken from the NAAL website http://nces.ed.gov/naal/estimates/index.aspx) shows 1992 and 2003 estimates and credible intervals for Connecticut adults lacking *Basic* prose literacy. In 2003, the estimated percentage of the adult population lacking *Basic* prose literacy stands at 9 percent with a credible interval width of 7. This means that there is a 95 percent chance that the true percentage of adults in Connecticut lacking *Basic* prose literacy would fall between 5.5 percent and 12.5 percent. For 1992, the estimate is 14 percent, with a credible interval width of 11.8. For that year, the true percentage would fall between 8.3 percent and 20.1 percent. Though the estimate for 2003 is lower than for 1992, the difference is not statistically significant.

Table 8.   Indirect estimate of adults in Connecticut lacking *Basic* prose literacy: 1992 and 2003

Location	Population size	Percent lacking *Basic* prose literacy	95 percent credible interval	
			Lower bound	Upper bound
Connecticut (1992)	2,590,405	14	8.3	20.1
Connecticut (2003)	2,668,989	9	5.5	12.5

SOURCE: NAAL State & County Estimates of Low Literacy web tool at http://nces.ed.gov/naal/estimates/index.aspx.

To investigate why the percentage of adults lacking *Basic* prose literacy in Connecticut is lower than the nation in 2003, we can obtain the demographic characteristics that are used as predictor variables in the 2003 statistical models for the United States and Connecticut, using the American FactFinder web tool as in the California example. Table 9 reveals that, in Connecticut's case, the low percentage of Blacks and Hispanics (used in the model as a single variable) and the population below 150 percent poverty line is most at variance with the U.S. figures, which helps us to understand why Connecticut has a lower estimated percentage lacking *Basic* prose literacy skills than the nation.

Table 9.   Comparison of predictor variables for Connecticut and the United States: Census 2000

Predictor variable	Connecticut	United States
Percent of adults ages 25+ with a high school education or less	44.5	48.2
Percent of Blacks/Hispanics	18.0	24.5
Percent of population below 150 percent poverty line	13.3	20.9
Percent of foreign-born people who stayed in the United States 0-20 years	6.6	7.7

SOURCE: U.S. Bureau of the Census, American FactFinder, Census 2000.

To understand why the percentage of adults lacking *Basic* prose literacy in Connecticut has dropped (albeit, not significantly) since 1992, we can look at the demographic characteristics used as predictors in both the 2003 and 1992 statistical models. Table 10 shows a reduction in the percentage of people with a high school education or less and an increase in the percentage of Blacks and Hispanics in Connecticut from 1990 to 2000. So more people in Connecticut completed high school and went on to postsecondary education despite the increase in the groups of people who are commonly associated with lower literacy rates. The increase (3.8 percentage points) in Blacks and Hispanics is not as big as the reduction (5.8 percentage points) in the people with a high school education or less, which gives some understanding of why fewer lack *Basic* prose literacy skills.

Table 10.   Comparison of predictor variables for Connecticut: 1990 and 2000

Demographics	1990	2000
Percent of adults ages 25+ with a high school education or less	50.3	44.5
Percent of Blacks/Hispanics	14.2	18.0

SOURCE: U.S. Bureau of the Census, American FactFinder, Census 2000 and 1990.

.

Discussion

The release in January 2009 of the small area estimates of low literacy brought much attention from state and local area administrators. Certainly a finding was the thirst for data to help support adult literacy efforts. This level of interest brought to light the need to educate the user community on how and when to use the model-based estimates, and created a challenge to communicate the complicated model-based estimates to the data users. In response, this report communicates both the strengths and limitations of the data-driven and model-based approach to making small area estimates of adult literacy. The following paragraphs provide some guidance in this regard, and further discuss the relevance of the findings for those who make policy decisions on adult literacy programs below the national level.

Users and Usages of the Findings

Decisions about adult education policies, practices, and funding allocation in states and counties are driven by information. State directors of adult literacy, governors, state commissioners of adult education, chief state school officers, adult and family literacy providers, program developers, literacy educators, representatives from the business community, state health care providers, state workforce development agencies, as well as state correctional education agencies, among other decisionmakers, can benefit from the NAAL, which provides them with model-based local area estimates for the first time since work by Reder (1997), along with previously unavailable reliable measures of the precision of those estimates (see footnote 2).

The allocation of government and regional funds often relies on state and county estimates for adults since a range of special services may be needed in areas with concentrations of the least literate adults.

For example, the allocation of government and regional funds often relies on state and county estimates for adults, since a range of special services may be needed in areas with concentrations of the least literate adults. In addition to arguing for new initiatives or policies and supporting budget requests to expand opportunities for career success in the global marketplace and quality of living for their adult populations, there are other examples of potential usage of state and county estimates provided by the NAAL. These include: tracking changes over time and thus gauging the effectiveness of state and local adult educational programs and policies; assessing adult literacy service needs and allocating funds appropriately; helping raise awareness outside the field to garner support for literacy programs; aiding researchers in exploring the relationships between literacy levels and other data related to income, age, gender, immigration, unemployment, health, and occupation, and to help determine why some geographic areas have higher rates of literacy than others; providing teachers with a better understanding of adult literacy demographics in their regions; and, finally, helping low literacy adults feel less isolated while receiving literacy services, knowing that many others share their predicament.

Limitations of the NAAL Indirect Estimates

Most indirect estimates at the county level are not precise because they have been generated by a statistical model, and not through a direct survey. This imprecision is reflected in the widths of the credible intervals, which are wide for some states and especially for many counties. Even though the model reduced the uncertainty of the county-level direct estimates, they are still fairly imprecise estimates.[10] The width of each credible interval is based on three factors: sample size, the strength of the relationship between predictor variables and the rate of low literacy individuals, and the size of the indirect estimates.

The first factor is the *sample size*. In general, smaller states and counties have fewer people and, therefore, smaller sample sizes, which means fewer people who were directly assessed among the states and counties sampled. The smaller the sample, the wider the credible intervals are. However, because the sample was not proportional to the size of county population, two counties of equivalent size may have different-sized samples.

The second factor is the *strength of the relationship between low prose literacy and the predictor variables* used in the model, including individuals who are foreign-born and in the United States for 0-20 years, have a high school education or less, are Black/Hispanic, and are below 150 percent of the poverty line. As a group, these variables explain 40 percent of the variance.
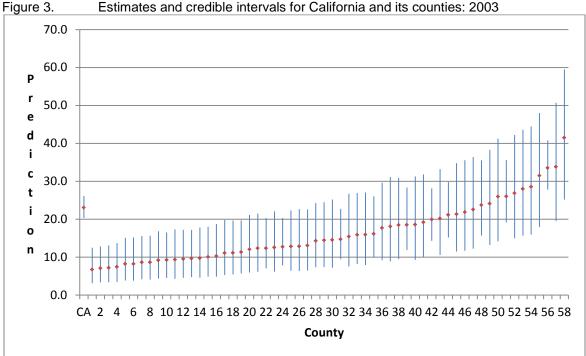
The third factor contributing to the width of the credible intervals is the *size of the indirect estimate*: the higher the indirect estimate, the wider the credible intervals. States range in their indirect estimates from 6 percent to 23 percent compared to the national direct estimate of 14.5 percent. These relationships can be seen in figure 3 for counties in California.

The figure shows the estimate (i.e., indirect estimate of the percentage of low literacy in prose) on the vertical axis, and counties on the horizontal axis, sorted first with the state on the far left and continuing with numbering from 1 to the number of counties in the state, sorted by the magnitude of their estimate (i.e., the percentage of adults lacking *Basic* prose literacy). The estimates, that is, indirect estimates, are shown as the dots in the graph. Their associated credible intervals are shown as the vertical lines. The indirect estimate and the credible interval for the state are on the far left side, showing an estimate of 23 percent with an interval of about 20 percent to 26 percent. The state estimate is a

---

[10] We computed the median confidence interval width to be 22 percent for the 264 counties with direct estimates. Using the confidence interval for the 264 counties with direct estimates as a substitute for credible intervals, we conclude that even though the model vastly reduced the uncertainty of the county-level direct estimates—decreasing from the estimate interval width from 22 percent (direct estimates) to 12 percent (indirect estimates), nonetheless, a 12 percent credible interval width is still a fairly imprecise prediction (Mohadjer et al. 2009, table 4-5).

weighted (according to population size) aggregate of the county-level estimates and therefore the state estimate falls in between the smallest and largest county-level estimates. Also we note that the interval width for the state is much smaller than for each of the counties due to the much larger sample size in the state in comparison to individual counties. The graph also shows that larger predicted values usually have wider credible intervals.

Figure 3.   Estimates and credible intervals for California and its counties: 2003

SOURCE: The NAAL State & County Estimates of Low Literacy web tool at http://nces.ed.gov/naal/estimates/index.aspx.

While the model-based estimates are imprecise, no other literacy assessment data are available for individual states and counties. The indirect estimates can provide state and county authorities with useful information when they judge the credible interval to be sufficiently small to support policymaking decisions (see California, Connecticut, and District of Columbia examples).

To obtain more precise state and county indirect estimates, it would be necessary to collect more direct data from individual states, collect data across more counties to obtain better direct estimates that could be used in modeling, and possibly obtain additional demographic variables. The latter would have to be consistent with Census variables to be useful in obtaining more precise indirect estimates in the future. Meanwhile, the indirect estimation model can be expanded in different directions as well. For example, projections can be created for Metropolitan Statistical Areas or for other sub-state areas of interest. Also, more research is necessary to develop a methodology for ranking states and counties, which does not now exist under the present model-based approach.

This page is intentionally blank.

References

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis.* 2nd edition. Texts in statistical science. Boca Raton, FL: Chapman and Hall/CRC.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., and Paulsen, C. (2006*). Literacy in Everyday Life: Results From the 2003 National Assessment of Adult Literacy* (NCES-2006-477). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Mohadjer, L., Kalton, G., Krenzke, T., Liu, B., Van de Kerckhove, W., Li, L., Sherman, D., Dillman, J., Rao, J., and White, S. (2009*). National Assessment of Adult Literacy: Indirect County and State Estimates of the Percentage of Adults at the Lowest Level of Literacy for 1992 and 2003* (NCES 2009-482). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

National Council of State Directors of Adult Education. (2006). *Adult Student Waiting List Survey, Adult Education State Directors.* Washington, DC: Author. Retrieved January 19, 2009, from http://www.ncsdae.org/.

Reder, S. (1997). *Synthetic Estimates of Literacy Proficiency for Small Census Areas. Prepared for the Division of Adult Education and Literacy, Office of Vocational and Adult Education, U.S. Department of Education*. Revised for internet publication. Accessed August 15, 2009, at https://www.casas.org/home/index.cfm?fuseaction=home.showContent&MapID=2800.

U.S. Department of Education. (2007). *Adult Education Annual Report to Congress Year 2004–05*. Washington, DC: Author, Office of Vocational and Adult Education.

White, S. (2008). *Assessing the Nation's Health Literacy: Key Concepts and Findings of the National Assessment of Adult Literacy (NAAL)*. Chicago: American Medical Association Foundation.

White, S., and Dillow, S. (2006). *Key Concepts and Features of the 2003 National Assessment of Adult Literacy* (NCES 2006-471). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

This page is intentionally blank.