Estimating literacy levels at a detailed regional level: An application using Dutch data

Maastricht University

Ineke Bijlsma,
Jan van den Brakel,
Rolf van der Velden, Jim Allen

Estimating literacy levels at a detailed regional level: An application using Dutch data

RM/17/018

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

Estimating literacy levels at a detailed regional level: An application using Dutch data*

Ineke Bijlsma[a], Jan van den Brakel[b], Rolf van der Velden[a] en Jim Allen[a]

a) Research Centre for Education and the Labour Market, Maastricht University
b) Statistics Netherlands and Maastricht University

Abstract

Policy measures to combat low literacy are often targeted at the level of municipalities or regions with an above-average population with low literacy levels. However, current surveys on literacy do not contain enough respondents at this level to allow for reliable estimates, at least when using only direct estimation techniques. To provide more reliable results at a detailed regional level, alternative methods must be used.

The aim of this paper is to obtain literacy estimates at the municipality level using model-based small area estimation techniques in a hierarchical Bayesian framework. To do so, we link Dutch Labour Force Survey data to the most recent literacy survey available, that of the Programme for the International Assessment of Adult Competencies (PIAAC). We estimate the average score, as well as the percentage of people with a low literacy level.

Additional complications arise, as the PIAAC framework assumes that test scores reflect an underlying latent construct. Moreover, as an adaptive design has been used with rotating modules, not all respondents are assigned the same test items. This is why an item response model is used with multiple imputation resulting in 10 so-called plausible values for the literacy proficiency level per respondent. Variance estimators for our small area predictions explicitly account for this imputation uncertainty.

The average literacy score is estimated with a unit-level model, while the percentage of low literates is estimated using an area-level model utilizing pooled variance. Optimal models are selected using a conditional Akaike information criterion score. Municipalities with less than 40,000 inhabitants were clustered with neighboring municipalities to ensure sufficiently large sample sizes.

The PIAAC survey is currently carried out in 36 countries. Most of these countries also have labor force surveys that contain similar information as the one used in this analysis. This opens up the possibility of applying the same method in other countries.

JEL classification: I26, J24

keywords: literacy, basic skills, municipality, region, small area estimation

* This study was made possible by a grant from the Reading & Writing Foundation. We thank Olivier Marie, Stefa Hirsch, participants of the SAE 2016 conference (17-19 August 2016, Maastricht) and the third PIAAC International Conference (6-8 November 2016, Madrid) for useful comments and suggestions.

1. Introduction

Research shows that cognitive skills play an important role in individual economic outcomes as well as in macroeconomic outcomes (Coulombe and Tremblay, 2007; Hanushek and Woessmann, 2008, 2011). People with high skill proficiency levels earn more, are more often employed, and generally face fewer economic disadvantages. Moreover, those with higher skills are more often engaged in civic and social activities (Organisation for Economic Co-operation and Development, or OECD, 2013a). Finally, looking at regional differences in skill levels, increases in overall levels of educational attainment seem to be associated with higher social returns in terms of wages (Moretti, 2004).

Generally, the skill levels in the Netherlands are among the highest in the world. In the Programme for the International Assessment of Adult Competencies (PIAAC) of 2012, the Netherlands ranked third in literacy, just behind Japan and Finland. Even so, there are still around 1.3 million people (11.9%) in the population of 16- to 65-year-olds who do not have the literacy skills necessary to function well in society (Buisman et al., 2013). The cost of low literacy in the Netherlands is estimated to be some 550 million euros per year (PricewaterhouseCoopers, 2013). With society's increased reliance on technology and digitalization, this group may face further marginalization.

Currently, a problem for policy makers is that detailed information on skill proficiency levels is not available at lower, regional levels, since most literacy surveys, including PIAAC, focus on the national level. However, policy measures are often targeted at municipalities or regions that are characterized by low literacy levels. For an efficient allocation of funding initiatives, local and regional governments need reliable data on the literacy levels in a particular municipality or region. Lack of these data restricts the efficient execution of a skills improvement strategy. It is therefore important to apply modern advanced estimation techniques to obtain accurate results at local and regional levels.

This paper focuses on refining PIAAC data into detailed regional results. With roughly 5,000 Dutch observations, direct estimators known from sampling theory can be applied to obtain reliable results at the national level. Most municipalities, however, have fewer than 20 observations. Direct estimators only use the observations from that domain and consequently have unacceptably large design variances. To increase the precision of municipal estimates, model-based small area estimation (SAE) techniques are applied in this paper. These methods assume an explicit statistical model to increase the effective sample size of each separate domain (for an overview, see Rao and Molina, 2015).

The basic idea of this regression method is that we assume that our dependent variable, literacy, is closely linked to personal characteristics such as age, gender, education, and labor status, which are also available in large survey based or administrative datasets. We also make the necessary assumption that the way these characteristics are linked is similar at both the national and detailed regional levels. Therefore, with detailed information for these characteristics at the regional level, it is possible to make more accurate model-based literacy predictions per municipality: a synthetic estimate. Unexplained variation between the domains is modeled with a random component in a multilevel model.

Model-based small area predictors can be expressed as the weighted average between the direct estimates based on PIAAC data and the aforementioned synthetic estimates, where the weights are based on the accuracy measures of the two estimators. If the underlying assumptions hold, this allows us to greatly reduce the variance of the estimates while introducing only limited bias to the estimates. Our goal is to generate usable estimations at the municipality level such that they can be used in policy decisions regarding the improvement of literacy proficiency levels.

SAE techniques are widely applied in social and economic sciences to produce reliable statistical information in detailed breakdowns. The World Bank, for example, applies a synthetic estimation procedure proposed by Elbers, Lanjouw, and Lanjouw (2003) to estimate poverty and income inequality in developing countries. The U.S. Bureau of the Census applies an SAE approach based on the method of Fay and Herriot (1979) to estimate income at low regional levels. These estimates are used to determine fund allocations to local government units. The National Research Council (2000) also used the Fay and Herriot (1979) model to produce county estimates of poor school-aged children in the United States for the allocation of supporting funds. Finally, Statistics Netherlands applies times series SAE methods to calculate official monthly unemployment figures (van den Brakel and Krieg, 2015).

To the best of our knowledge, SAE techniques in the context of literacy skills have only been applied in two earlier papers, both of which take a quite different approach than the one we present here. Gibson and Hewson (2012) use U.K. census data and SAE modeling to obtain synthetic estimates of literacy levels in geographical areas (the 6,781 so-called middle layer super output areas in England, as well as their aggregates). Although the initial input is based on the 2011 Skills for Life Survey, the output for regional units is solely based on synthetic estimates. Moreover, the authors use microsimulation to produce the likely number of people in each area with certain combinations of characteristics. Another caveat of their data is that the census data are from 2001 and thus 10 years before the 2011 Skills for Life Survey was administered. Yamamoto (2014) adopts a more similar approach to ours, also using PIAAC and labor force survey (LFS) data. However, the author's aim is to only produce synthetic estimates. Yamamoto's paper is based on Canadian data, which comprised a much larger group (over 20,000 respondents) in the PIAAC sample than the other countries, thus allowing for more regional accuracy. The author showed that the synthetic estimates for provinces based on the LFS data are quite similar to the direct estimates for these provinces based on PIAAC data and concluded that this is an indication that synthetic estimates can be used for the years between subsequent PIAAC data collections.

The contribution of this paper is the application of SAE techniques to estimate municipalities' literacy levels that are a weighted average of direct and synthetic estimates, with the weights based on the uncertainty measures of both estimates. This approach has the advantage that, in large municipalities with relatively large sample sizes, the direct estimates make a relatively large contribution to the final estimate whereas, in small municipalities, the final estimate is dominated by the synthetic estimator. The PIAAC data setup presents a number of challenges which prevents straightforward estimations. Addressing these challenges is novel in the application of SAE techniques. Respondents were randomly assigned to (parts of) the literacy tests. This requires imputation techniques to account for missing observations. Moreover, the PIAAC tests follow an adaptive design, so that respondents are assigned items that are close to their expected proficiency levels, based on the scores of previous questions. The model follows an item response theory (IRT) approach, which assumes that the scores on the tests are based on a latent construct that cannot be measured directly. Instead, for each respondent, 10 plausible values are calculated and several replicate weights are constructed, which can be seen as a form of multiple imputation. This approach allows for the construction of point estimates as well as variance estimates for literacy. We use both a unit-level model (Battese, Harter, and Fuller, 1988) and an area-level model (Fay and Herriot, 1979) and detail how to incorporate this structure into our SAE approach.

Our paper is organized as follows. Section 2 covers the definition of literacy, as well as the data description. Section 3 details the techniques of the small area predictors for this application. Section 4 presents the models finally selected and their fit. Section 5 evaluates the model and presents robustness checks. Section 6 reports the results of our analysis and Section 7 concludes the paper.

2. Definition of Literacy and Data Description

In this section, we define the concept of literacy and describe the measures of literacy we wish to use and the data needed for our analysis. These specifications are at the basis of our model.

PIAAC

The data set we are using to measure literacy skills is from the 2012 PIAAC survey (OECD, 2013a), an international survey led by the OECD. It is designed to map skills and competencies in developed countries, measuring the numeracy, literacy, and problem solving skills of adults. In addition, it collects a range of information on the way the respondents use these skills and how often.

Literacy in PIAAC is defined as "the ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and develop one's knowledge and potential" (OECD, 2013a, p. 59). It does not include the ability to write or produce texts but focuses on the ability of an individual to interact with written text. It is this definition that will be used throughout the paper.

Data collection in the Netherlands took place from August 1, 2011, to March 31, 2012, and was undertaken in the respondents' homes. The target population was between 16 and 65 years of age, residing in the country at the time the data were collected. For the Netherlands, 5,170 respondents randomly were selected by one-stage stratified simple random sampling without replacement from the Dutch population register. Strata were formed by municipalities. The sample weights are based on the sampling design.

The PIAAC survey used two data collection modes to measure skill proficiency levels: a paper and pencil delivered mode and a computer-delivered mode. Respondents with no computer experience or who refused to take the test on the computer were routed to the paper and pencil mode, as well as those who performed poorly on the computer based core section taken before the test.

Both modes started off with a core assessment of basic numeracy and literacy questions. The small proportion of respondents that performed poorly was routed directly to the reading component skill measures.

After the core section, the respondents in the paper and pencil delivered mode were randomly assigned to a cluster of either 20 literacy or 20 numeracy questions, followed by reading component skill measures. In the computer-delivered mode, the respondents who performed well were routed at random to one of three possible outcomes: 50% were assigned both literacy and numeracy tasks, 33% were assigned problem solving tasks combined with either literacy or numeracy tasks, and 17% were assigned only problem solving tasks.

For literacy, the questions differed in content, cognitive strategies, and context. A multistage adaptive design was used between the items and an algorithm determined the next item depending on the responses.

This survey design was such that different groups of respondents were routed to items with potentially various degrees of difficulty, disallowing direct comparisons between the respondents' test scores. Therefore, the item responses were first fitted to an IRT model. After item calibration, the IRT model was combined with a latent regression model using information from the background questionnaire in a population model to further improve accuracy. From this step, 10 plausible values were drawn on a scale of zero to 500. Lastly, a replication approach (Johnson and Rust, 1992) was used to estimate the sampling variability as well as the imputation variance associated with the

plausible values. Note that, for respondents not routed to one of the literacy components, the plausible values were imputed.

Variance estimation, taking into account the sample design, the selection process, the weighting adjustment, and the measurement error through imputation, is carried out using a replication approach. For the Netherlands, a paired jackknife estimator was used with 80 replicate weights. To take this survey design into account, we used the PIAACTOOLS package of Pokropek and Jakubowski (2013). A detailed description of the construction of the variance term can be found in the PIAAC *Technical Report* (OECD, 2013b).

Literacy scores can then be categorized at multiple levels based on the scoring range. For example, level 1 literacy consists of a score between 176 and 226 points. At level 1, one can complete simple forms, understand basic vocabulary, and read continuous texts but would have trouble making low-level inferences. In comparison, the highest level of literacy, level 5, is determined by a score of 376 or higher. These levels are described in detail in the PIAAC *Reader's Companion* (OECD, 2013c).

We classify someone as having a *low literacy level* when that individual has literacy level 1 or below. This means we include all people who have trouble understanding and carrying out relatively basic literacy tasks.

There are various options for describing the literacy levels in a region. One straightforward method would be to look at the average literacy test score. This is a good way of providing a quick snapshot of the literacy level within the whole population, which uses the full information available in the PIAAC data. A limitation of the average score is that it provides no further information as to how literacy levels are distributed within regions.

Another measure would be to give an estimate of those who would be classified as having a low literacy level based on the earlier definition. This measure would be most important for policy making, since this is the group that would benefit the most from policy interventions. A disadvantage of this measure is that information is lost due its dichotomous nature.

Taken together, both measures—the average score and the proportion of those with low literacy scores—provide the best picture of the situation concerning literacy levels in a region. Generally speaking, the average score can be interpreted as an indicator of the general literacy level within an area. As can be seen in the results, relatively large cities, especially those with universities, generally do well in this indicator, since there is a strong concentration of highly educated people in these areas. Interestingly, however, when we look at the percentage of low literates, larger cities also do much worse, since they usually contain higher percentages of immigrants and other low-skilled workers. Conversely, some regions with an average overall literacy score consist primarily of people with scores that are quite close to the national average and may therefore have a below-average percentage of low literates.

As stated in the introduction, the Netherlands has a relatively high average literacy score of 284. The percentage of low literates is relatively low, at 11.9%. The total number of respondents in PIAAC is 5,170, but for some respondents the municipality is unknown. We are left with 5,073 respondents, which leads to slight changes in the estimates (see Table 1).

Table 1: Summary of the statistics of the target sample (PIAAC)

	Mean	St. Error	Lower Bound	Upper Bound
Average Score	283.94	0.68	282.61	285.27
Low Literacy	12.00	0.46	11.07	12.86

LFS

SAE requires auxiliary data that include personal characteristics which are closely linked to literacy levels. For our research, we have decided to use the Dutch LFS. The LFS is a continuous survey that has been carried out by Statistics Netherlands since 1987. The goal is to study the relations of people's characteristics with their (future) position in the labor market. The LFS's features (large sample sizes, good overlap in questions about personal characteristics) make it a good choice for auxiliary data.

In our selected timeframe, interviews for the LFS took place face to face and by phone. The weights are calculated in two steps using general regression estimators (Särndal et al., 1992). In the first step, design weights are derived from the sample design and account for differences in selection probabilities. In a second step, the design weights are calibrated to available auxiliary information for which the true population distributions are known from registrations to correct, at least partially, for selective non-response.

To ensure sufficient data from each area, we chose to include three years of LFS data. To be precise, we chose 2010–2012 as our sample period for two reasons: The first reason is that the PIAAC survey ran from August 2011 until May 2012, so we would like to adhere as closely as possible to this period. Second, from 2013 onward, changes were made in the LFS questionnaire, so, for comparison purposes, it was better to avoid possible mismatches in the data. We apply the same age restriction (between 16 and 65 years old) as in the PIAAC survey.

Since the LFS has a rotating panel design, people were asked multiple times to participate and thus are included multiple times. We weight these people over the number of samples within our selection, so that they are counted only once and not multiple times. This leaves us with about 890,000 observations obtained from 309,000 respondents.

3. SAE

Due to time and budget limitations, most social, demographic, and economic analyses are based on a sample of individuals instead of a complete enumeration of the entire population. Therefore, surveys are usually set up to be conducted over a representative subset of the population obtained from a probability sample and the results are extrapolated over the entire population using direct estimators. While this setup allows for accurate statistics for larger areas, the sample size for smaller subpopulations is frequently simply too small to create reliable estimates based on direct estimators. Sample size is restricted by available resources and time and, in many surveys, it is too costly to sample a large number of individuals within each subpopulation of interest.

SAE is a method specifically developed for this kind of problem. SAE augments the information contained in the primary data source by using auxiliary covariate data at the level of smaller units or subpopulations—the so-called small areas of interest—contained in register data sets or large surveys, such as the LFS. By means of these auxiliary data, we can compute covariates per area that are related to the variable of interest. If these covariates have good predictive power, the variable of interest can be estimated per area and these estimates can be used together with the original survey data to produce predictions that are more precise compared to the direct estimates.

A large amount of SAE procedures are available in the literature. See Rao and Molina (2015) for a detailed overview or Pfeffermann (2013) for a more summarized overview. In this paper, we have chosen a multilevel modeling approach. The models are fitted in a *hierarchical Bayesian* (HB) framework. All models, including the model selection measures, were run using the fSAE function in the software program R, available via the hbsae package (v. 1.0, available in the Comprehensive R Archive Network; Boonstra, 2015).

It is important to keep some things in mind when interpreting the results from SAE. Most importantly, the estimates are not from direct data. There are various possible biases that could lead to outcomes that differ from those if all the results were based on direct estimates of a larger sample. One important possible bias is due to the assumption that the relations between literacy and personal characteristics at the national level are the same at the regional level. Violation of this assumption can lead to large differences between the regional estimations and "true estimates."

Literacy Measures

As stated earlier, we are interested in two measures of literacy per area: the average score and the percentage of low literates. In the first case, the dependent variable *y* is continuous per individual and area and we assume that *y* has a linear relation with the chosen covariates *X*. In this case, we use the basic unit-level model originally proposed by Battese, Harter, and Fuller (1988), where the input variables for the model are individual measurements obtained from the sampling units. We go into more detail in the section below on the unit-level model.

However, in the second case of the percentage of low literates, the dependent variable is dichotomous at the individual level, since each plausible value will be binary, equal to one if the score is below the low-literacy cutoff point of 226 and zero otherwise. We decided to model the percentage of low literates with a basic area-level model, as originally proposed by Fay and Herriot (1979). In the next two sections, we elaborate both the area-level model and the unit-level model. A complete overview can be found in the work of Rao and Molina (2015). Afterward, we explain how we incorporated the PIAAC imputation structure in the estimations.

Area-Level Model

The input for the area-level model is provided by the direct estimates for the domains. Let $y_{i,a}$ denote the average of the 10 plausible values of an individual *i* who belongs to municipality *a*, as observed in the original survey data (PIAAC). These average values are used to construct the area average score of literacy, for example, $\bar{y}_a$, using the paired jackknife estimator (see also Section 2). The jackknife is also used to estimate the variance of $\bar{y}_a$, denoted ψ_a^2, and accounts for sampling error, the uncertainty of multiple imputation for missing values, and the uncertainty of the IRT model underlying the adaptive tests for literacy, using both replicate weights and plausible values. Furthermore, let $\bar{X}_a$ denote the vector with the population means of the auxiliary variables derived from the LFS used for calibration. The sample area means for the auxiliary variables derived from the PIAAC sample are denoted $\bar{x}_a$. Survey errors regarding the estimation of $\bar{X}_a$ from the LFS are assumed to be negligible and are not taken into account.

For each area, we can construct the survey regression estimator $\hat{y}_a^{surv}$ as

$$\hat{y}_a^{surv} = \bar{y}_a + (\bar{X}_a - \bar{x}_a)^t \beta,$$

where β is the vector with regression coefficients from the linear model that describes the relation between the target variable *y* and the auxiliary variables *x*. Then, we assume that the survey regression estimator can be written as the true mean plus a sampling error term, such that

$$\hat{y}_a^{surv} = \mu_a + e_a, \quad (1)$$

where μ_a is the true area mean and e_a is an independently distributed sampling error e_a that has expected value zero and sampling variance ψ_a^2. In turn, we model the true mean with a random effect model,

$$\mu_a = \alpha + \bar{X}_a \beta + u_a, \quad (2)$$

where α is the intercept, $\bar{X}_a$ the area covariate averages, β the vector of coefficients of covariates, and u_a a random effect to take into account area-level variation not explained by the fixed part of the equation. The random effects are assumed to be normally and independently distributed, with an expected value equal to zero and model variance σ^2. This model is called the Fay–Herriot model.

Inserting (2) into (1) gives the measurement equation

$$\hat{y}_a = \alpha + \bar{X}_a \beta + u_a + e_a. \quad (3)$$

Based on this model, the best linear unbiased predictor (BLUP) estimator for the domain means is equal to (Rao and Molina, 2015, Sections 7.1.1 and 7.1.2)

$$\hat{y}_a^{BLUP} = \varphi_a(\hat{y}_a^{surv}) + (1 - \varphi_a)(\bar{X}_a^t \hat{\beta}), \quad (4)$$

where $\hat{\beta}$ is the vector of fixed effects estimated at the national level and φ_a is a weight between the direct and synthetic estimator given by

$$\varphi_a = \sigma^2/(\psi_a^2 + \sigma^2).$$

Now, if in (4), the variance of the random area effects σ^2 is replaced by its estimator $\hat{\sigma}^2$, the empirical BLUP (EBLUP) estimator is obtained (Rao and Molina, 2015, Sections 7.1.1 and 7.1.2). Moreover, the sampling variance ψ_a^2 is assumed to be known; however, in practice, this is not true and, in this application, it is replaced by its estimator obtained with the paired jackknife. The mean squared error (MSE) of the EBLUP accounts for the additional uncertainty that is introduced, since σ^2 is replaced by its estimator $\hat{\sigma}^2$ but ignores the uncertainty of using an estimator for ψ_a^2, which is common practice in SAE procedures. If $\hat{\sigma}$, the estimated model variance, is large, more weight is shifted to the direct component. If ψ_a^2, the sampling variance (taken to be the survey variance) is large, more weight is shifted to the synthetic component.

Maximum likelihood (ML) procedures are often applied to estimate the model variance σ^2. Particularly in the case of strong auxiliary information, the ML estimate for σ^2 tends to zero. This is undesirable because, in this situation, the EBLUP estimator gives full weight to the synthetic estimator and ignores the sample information, even in larger domains with a substantial sample size. These problems are circumvented in this paper by using an HB approach, as outlined by Rao and Molina (2015, Section 10.3).

The HB model is based on (3) under the assumption that $e_a \sim N(0, \psi_a^2)$ and $u_a \sim N(0, \sigma^2)$. For β and σ^2, a flat prior distribution is assumed. The HB estimates for the area means, including their MSEs, are obtained by the posterior means and posterior variances of the posterior density for the domain means u_a. These estimates can be evaluated using separate one-dimensional numerical integrations. As shown by Rao and Molina (2015, Section 10.3), the estimator for σ^2 is always unique and positive and results in more plausible estimates than the EBLUP estimator.

Due to the small sample sizes, the estimates for the variances of the survey regression estimates will be unstable. Under the assumption that each municipality has equal population variances, we decided to use a pooled variance estimator for each area to increase precision. The variance approximations obtained with the jackknife are pooled using an analysis of variance type pooled estimator:

$$\psi_a^{2;P} = \frac{1}{N_a} \frac{\sum_{a=1}^{m}(N_a-1)\psi_a^2}{\sum_{a=1}^{m}(N_a-1)},$$

where *m* is equal to the total number of areas. Furthermore, it was clear that some municipalities were scoring unnaturally low (one was even negative) and were underestimated due to the linearity of the model. Therefore, two post-result changes were implemented. First, we acknowledged that the model had problems estimating the true percentages in areas where the percentage of low literates is very small (< 5%), which is further considered in the results; so, during categorization, we marked these municipalities as having a very small percentage (0–5%) of low literates and grouped them together when publishing the results. Second, a choice was made to benchmark the results such that they would add up to the national level, minimally adjusting the results by means of the number of samples and the MSE per area, using a Lagrange function.

Unit-Level Model

As before, let $y_{i,a}$ denote the average of the 10 plausible values in the literature regarding an individual *i* who belongs to municipality *a*, observed in the original survey data (PIAAC). The expected value of the true mean is then equal to

$$y_{ia} = \mu_{ia} + e_{ia} = \alpha + x_{ia}^t\beta + u_a + e_{ia}, \quad (5)$$

where x_{ia} is a vector with covariates for respondent *i* from area *a* and u_a is an area-specific random effect assumed to be independent and identically distributed. We assume e_{ia} is a measurement error for respondent *i*, where $\hat{e}_{ia}$ has expected value zero and variance σ_e^2. The EBLUP estimator is then equal to

$$\hat{y}_a^{BLUP} = \varphi_a\,(\hat{y}_a^{surv}) + (1-\varphi_a)\,(\bar{X}_a^t\hat{\beta}),$$

where the weight φ_a, dependent on area size N_a, is given by

$$\varphi_a = \sigma^2/(\sigma^2 + \sigma_e^2/N_a).$$

The HB model is obtained with (5) with the assumption that $e_{ia} \sim N(0, \sigma_e^2)$ and $u_a \sim N(0, \sigma^2)$. Furthermore, flat priors are assumed for β, σ_e^2, and σ^2. The HB predictors for the area means, for example, $\hat{y}_a^{HB}$, with their MSEs are computed as the posterior means and posterior variance of the posterior distribution of μ_a in a similar way as for the area-level model. The resulting integrals are solved using numerical integration.

Unlike the area-level model for the percentage of low literates, where the imputation uncertainty is taken into account when constructing $\bar{y}_a$, the unit-level model as described above does not take into account the imputation uncertainty.

Multiple imputation is one way to take into account this imputation uncertainty, as shown by Rubin (1996). The plausible values generated with the PIAAC software are used to calculate multiple HB predictions for the areas. Let $\hat{y}_{aj}^{HB}$ denote the HB prediction for area *a* based on the *j*th set of

plausible values generated for the PIAAC sample and $MSE(\hat{y}_{aj}^{HB})$ denote the posterior variance of $\hat{y}_{aj}^{HB}$. The final HB prediction for area *a* is defined as

$$\hat{y}_a^{imp} = \sum_{j=1}^{k} \frac{\hat{y}_{aj}^{HB}}{k},$$

where *k* is the total number of plausible values. The associated variance–covariance matrix V^{imp} is equal to

$$V^{imp} = W + \frac{k+1}{k} B,$$

where the within-imputation variability *W* is obtained as the mean over the MSE of the HB small area predictions:

$$\mathrm{W} = \sum_{j=1}^{k} \frac{MSE(\hat{y}_{aj}^{HB})}{k}.$$

The between-imputation variability *B* is

$$B = \sum_{j=1}^{k} \frac{(\hat{y}_{aj}^{HB} - \hat{y}_a^{imp})^2}{k-1}.$$

4. Model Fitting

Merging of Municipalities

In 2012, the Netherlands comprised 415 municipalities. However, some municipalities are quite small and we cannot guarantee that their LFS data cover enough respondents to provide an accurate representation of its inhabitants. Therefore, it is necessary to work with municipality clusters instead. We use 40,000 as the minimum number of residents per area to ensure the LFS estimates can be considered reliable. Municipalities with fewer residents are clustered together with adjacent municipalities. During this merging, we made sure that all the areas could still be nested in larger official area aggregates, such as provinces or so-called COROP regions. The latter refer to a Statistics Netherlands 40-area classification based on educational provisions. Finally, 208 municipality clusters are obtained for which small area estimates about literacy will be made. In the PIAAC sample, the minimum number of observations for these clusters is six, the maximum is 146, and the median is 20.

Variable Selection

The variables available from the LFS are considered for use as auxiliary variables at the area level and in the unit-level model for literacy. A list of potential auxiliary variables and descriptive results are listed in Table A1 in the Appendix. Optimal models are selected by means of the conditional Akaike information criterion (cAIC) using a stepwise backward variable selection procedure.

Covariates were removed one by one until a minimum for the cAIC was reached for the unit-level model on literacy scores. The list of the remaining predictors is as follows:

- Age, Age squared
- Immigrant Status (non-immigrant, first-generation immigrant, second-generation immigrant)
- Years of Schooling (based on the highest level of education completed)

- Area of Study (eight categories) of the highest level of education completed
- Vocational Education: Dummy based on the highest level of education completed
- Employment Status (Currently in education, Self-employed, Full-time working employee, Part-time working employee, Other)
- Occupational Status Measure (ISEI-08, set to the average for the non-working population)
- Interaction terms of Immigrant Status with Age
- Interaction terms of Gender, Full-time working employee, Part-time working employee, Currently in education, and Vocational Education with Years of Schooling

For the area-level model on the percentage of the low literates, almost the same model was selected. An improved model could be constructed by leaving out the self-employed and one dummy regarding the area of study (ΔAIC = 2.9), leading to minor differences. The reason for using the same model for both literacy measures is that they are both measures of the literacy level in a municipality and therefore should be affected by the same set of predictors, at least theoretically. As a robustness check, we also applied the full model for the percentage of low literates and observed that estimates of this proportion per municipality using the two models are not substantively different.

Gender as a separate effect was discarded by the cAIC score and the results do not change when including it in the model. Furthermore, the interaction terms help with estimating effects of variables not captured in our data sets. For example, international knowledge workers would be classified as immigrants, which is generally a negative indicator. By including the interaction effect with years of schooling, we can partially correct for this.

Indicators in the Data

The weights in the PIAAC data have already been calibrated to a number of auxiliary variables whose population totals are known. However, since our LFS data period is slightly longer and covers a slightly different set of variables, it is prudent to check for any possible differences within the data. In Table A1 of the Appendix, we compare the weighted distribution of auxiliary variables between the two data sets to verify that there are no large discrepancies between the data sets with respect to the auxiliary variables. None of the variables are statistically significantly different from one another at the 5% level, aside from the percentage of second-generation immigrants.

5. Model Evaluation

The SAE results can differ from the direct results for a number of reasons. The most important reason is why the SAE techniques are applied in the first place, namely, to improve the precision of the direct municipality estimates. However, it is important to make sure the differences are not dominated by bias. Since SAE techniques explicitly rely on statistical models to improve the effective sample size in the separate domains, one must evaluate the underlying assumptions of the models. Model misspecification can easily result in heavily biased domain estimates. This section evaluates the normality assumptions underlying the applied models. Furthermore, direct domain estimates are compared with model-based small area predictions to assess possible bias. Finally, the improvement in precision is evaluated by comparing the standard errors of both estimators.

Robustness Checks

The direct estimates at the national level are precise and unbiased, since they do not depend on model assumptions and are based on a large sample. Therefore, the difference between the model-

based small area predictions, aggregated at the national level, with the direct estimates at the national level is often used as a measure of bias in SAE.

As noted earlier, in Section 3, benchmarking was applied to remove differences between model-based area estimates aggregated at the national level and direct estimates at the national level. Small area estimates for literacy scores and the percentage of low literates at the national level are obtained by calculating the mean over the municipalities weighted by the number of residents in 2012. Table 2 displays the results of the non-benchmarked estimates against the (robust) national results.

Table 2: Estimated aggregated results at higher levels, without benchmarking

Type	Direct	SAE (*)
Average Literacy	283.9	287.9
% Low Literacy	12.0%	12.8%

* indicates the average of the SAE results over municipalities, weighted by the number of residents in 2012.

For both measures of literacy, the SAE scores are a slightly overestimated. The average literacy is greater than the upper bound of 285.3 for the direct estimates given in Table 1. The percentage of low literates estimates are contained within the 95% confidence interval, but barely. On the basis of these results, we decided to benchmark our estimates.

After benchmarking, we look at the differences between the direct estimates and the SAE results. Two measures are applied to summarize the differences between the direct and model-based domain estimates. The first one is the *mean relative difference* (MRD), in percentages, defined as

$$MRD = \frac{1}{M}\sum_{a=1}^{M}\frac{(\hat{y}_a^{direct} - \hat{y}_a^{HB})}{\hat{y}_a^{direct}} * 100.$$

The second one is the *absolute mean relative difference* (AMRD), in percentages, defined as

$$AMRD = \frac{1}{M}\sum_{a=1}^{M}\frac{|\hat{y}_a^{direct} - \hat{y}_a^{HB}|}{\hat{y}_a^{direct}} * 100.$$

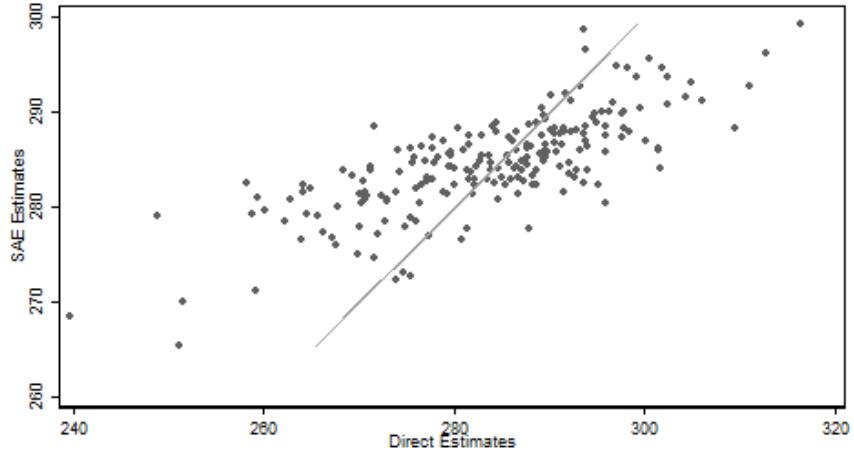
Table 3 gives the MRD and AMRD for the two literacy measures.

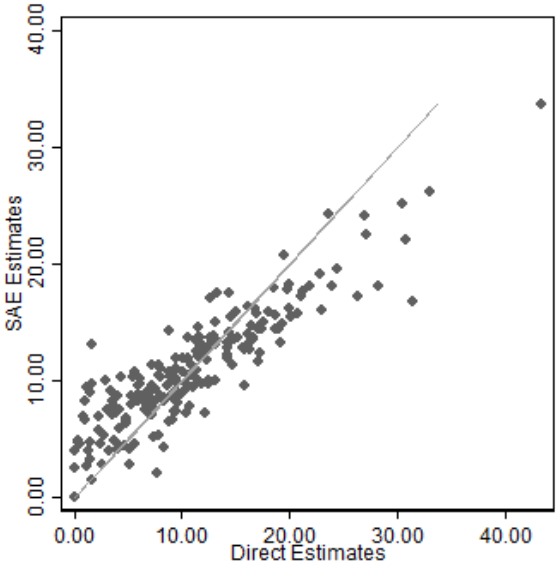
Table 3: Measures of quality of the estimates (%)

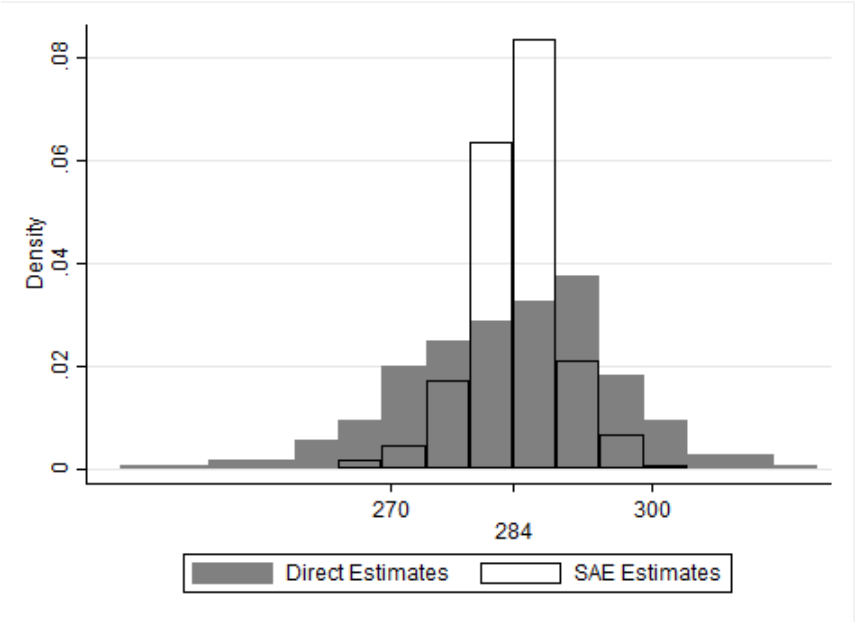
	Average Literacy	% Low Literates
MRD	-0.47	-0.52
AMRD	2.5	2.5

The MRD for both estimates is quite small, only half a percentage point. Since it is negative, the SAE estimators are generally slightly bigger. When we look at the absolute difference, we see a 2.5% mean difference for both estimators. To interpret these differences in more detail, we compare the distribution of the SAE estimates with the distribution of the direct results from PIAAC.

Figure 1: Histograms and distribution plots of the direct results and the SAE results (left, literacy scores; right, % low literates; the straight line is the diagonal).







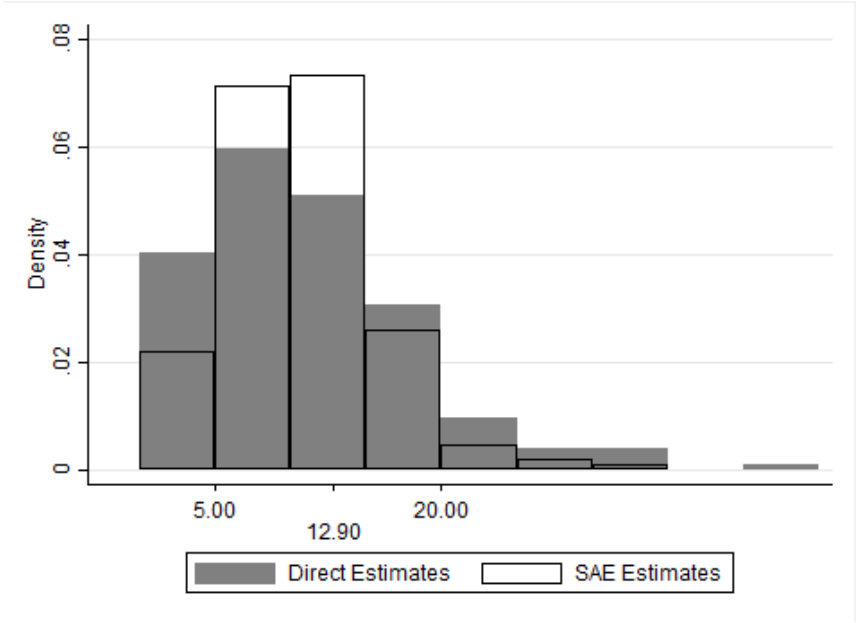


Figure 1 shows the tendency of the SAE estimates to trend toward the mean. Regarding the average literacy scores, the scores at the top consist mostly of those for university cities, where the number of students seems to be oversampled. The scores at the bottom are mostly for small villages.

For the percentage of low literates estimates, the top end of the distribution is close to the distribution of the direct estimates; however, note that the SAE results for the average and below-average percentage of low literates are higher than the direct results. The relatively high proportion of municipalities that perform well on the percentage of low literates (with percentages in the range of 0–5%) in the direct estimates could be due to the fact that these municipalities are very small and have few direct observations in PIAAC. Therefore, these differences would be a result of the improved accuracy of the point estimates.

Figure 2 shows the scatter plots of the fitted values of both SAE measures versus the quantiles of the residuals. No pattern can be distinguished within the two graphs, meaning the residuals are well behaved.

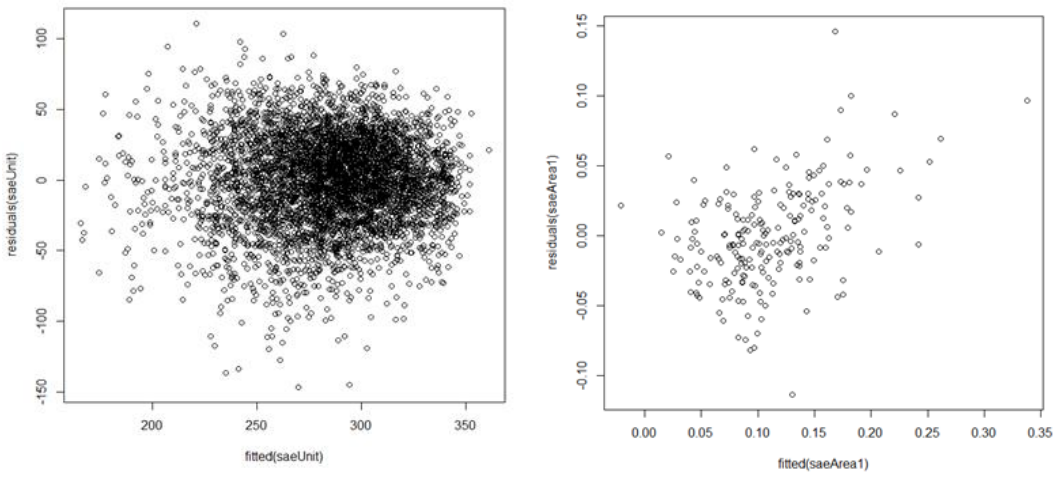


Figure 2: Fitted values versus the residuals of the unit-level estimates of the estimated literacy scores (left) and the area-level estimates of the percentage of low literates after benchmarking (right).

Finally, in Figure A1 of the Appendix, using Q-Q plots, we graphically check whether the SAE estimates, residuals, and random effects for both models all follow a normal distribution. If the points follow a linear line, this suggests the quantiles of both distributions line up and our estimates follow a normal distribution. While this is generally the case in our results, some discrepancy is visible at the tail ends. The most notable deviation is within the estimations of average literacy, where the values diverge a bit in the lower quantiles.

Reduction in Standard Error

To measure the increase in precision obtained with the SAE techniques, the *mean relative difference in standard errors* (MRDSE) is used, which is defined as

$$MRDSE = \frac{1}{M}\sum_{a=1}^{M}\frac{(SE_a^{direct} - SE_a^{HB})}{SE_a^{direct}} * 100.$$

The results are shown in Table 4. The MRDSE for average literacy is 68%, which, compared to the direct estimates, is a significant reduction. For the percentage of low literates, the reduction measure is 51% (31% when compared to the pooled variance) but, as a less powerful model, lower

returns are to be expected. In terms of the difference of estimates, both models show an absolute difference of about 2.5%, as follows from Table 3.

Table 4: Measures of the quality of estimates (%)

	Average Literacy	% Low Literates*
MRDSE	68	51 (31)

* indicates the numbers in parentheses are compared to the standard errors of the pooled variance instead of the direct standard errors.

In Figure 3, we look at the number of respondents in PIAAC versus the standard error of the direct estimates, as well as the SAE results for the average literacy scores per municipality. Given the high frequency of respondents numbering between five and 20 per municipality, we decided to graph these on a logarithmic scale.

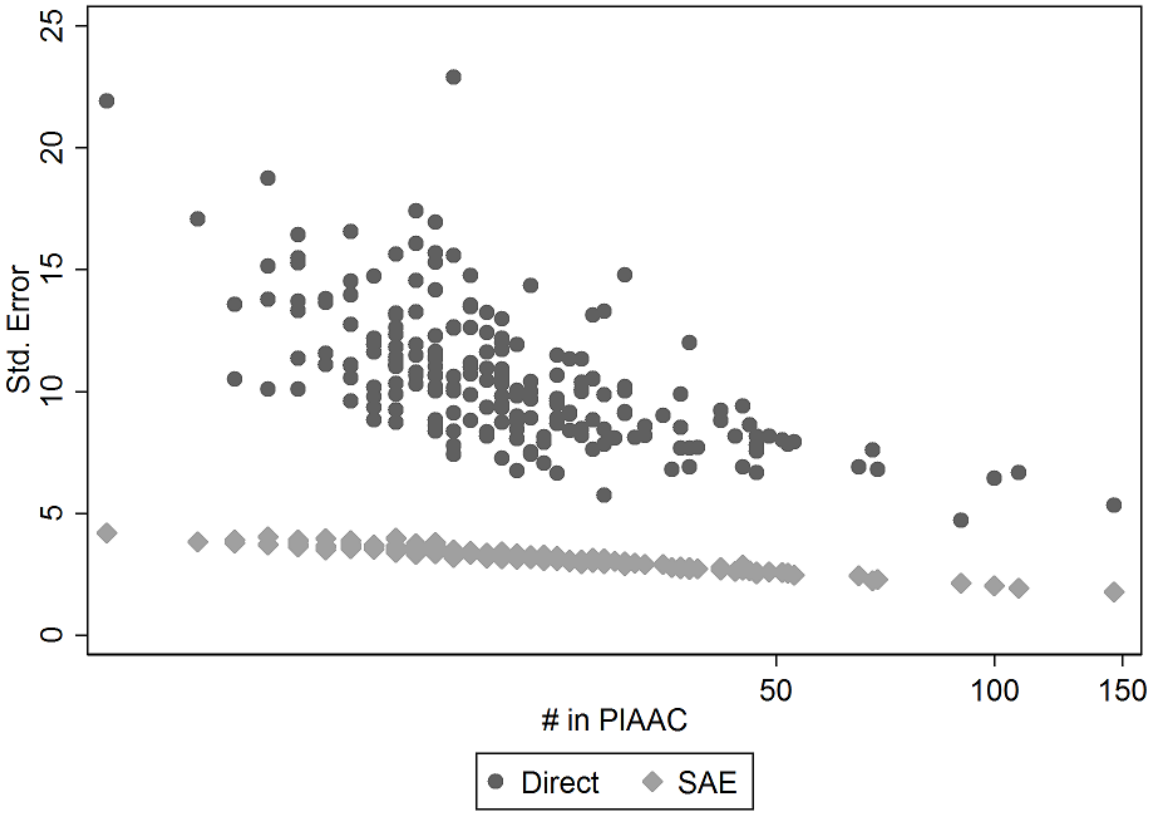


Figure 3: Standard errors versus the (logarithmic) number of PIAAC respondents for both the direct estimates and the SAE for the estimated literacy scores per municipality.

For small sample sizes, the SAE results show a large decrease in terms of standard errors compared to the direct estimator, whose margin of error is far too large when it comes to accurate point estimates. As the sample size increases, the difference between the two estimators decreases greatly.

In Figure 4, we look at the standard errors for the percentage of low literates. Here, the standard errors of the direct estimator are much more spread out and sometimes even zero (due to the direct estimator being zero). When compared to the pooled data, these are much closer to the SAE results, but there is still a significant gain in municipalities with low numbers of PIAAC respondents.

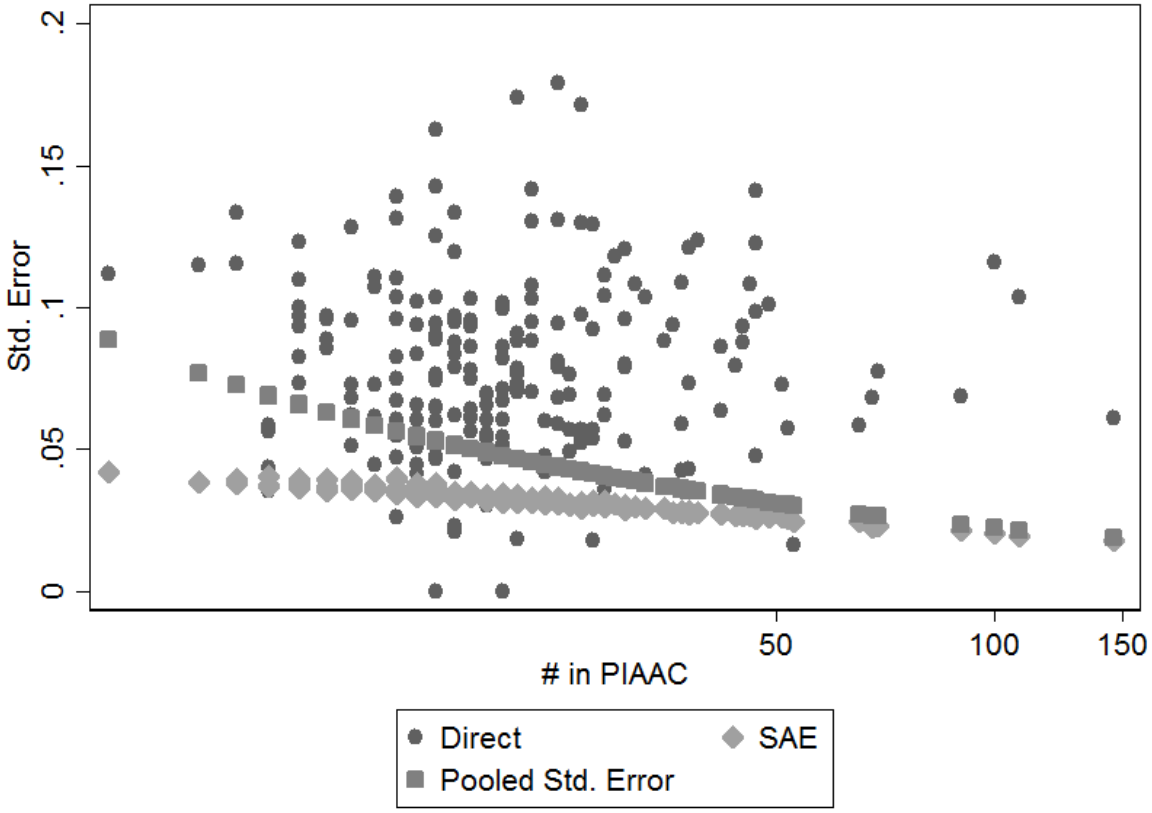


Figure 4: Standard errors versus the (logarithmic) number of in PIAAC respondents for both the direct estimates and the SAE for the percentage of low literates per municipality.

6. Results

In this section, we present the substantive results graphically, review them, and discuss the differences in results for the two chosen measures of literacy. The full list of results per municipality can be found in Table A2 of the Appendix.

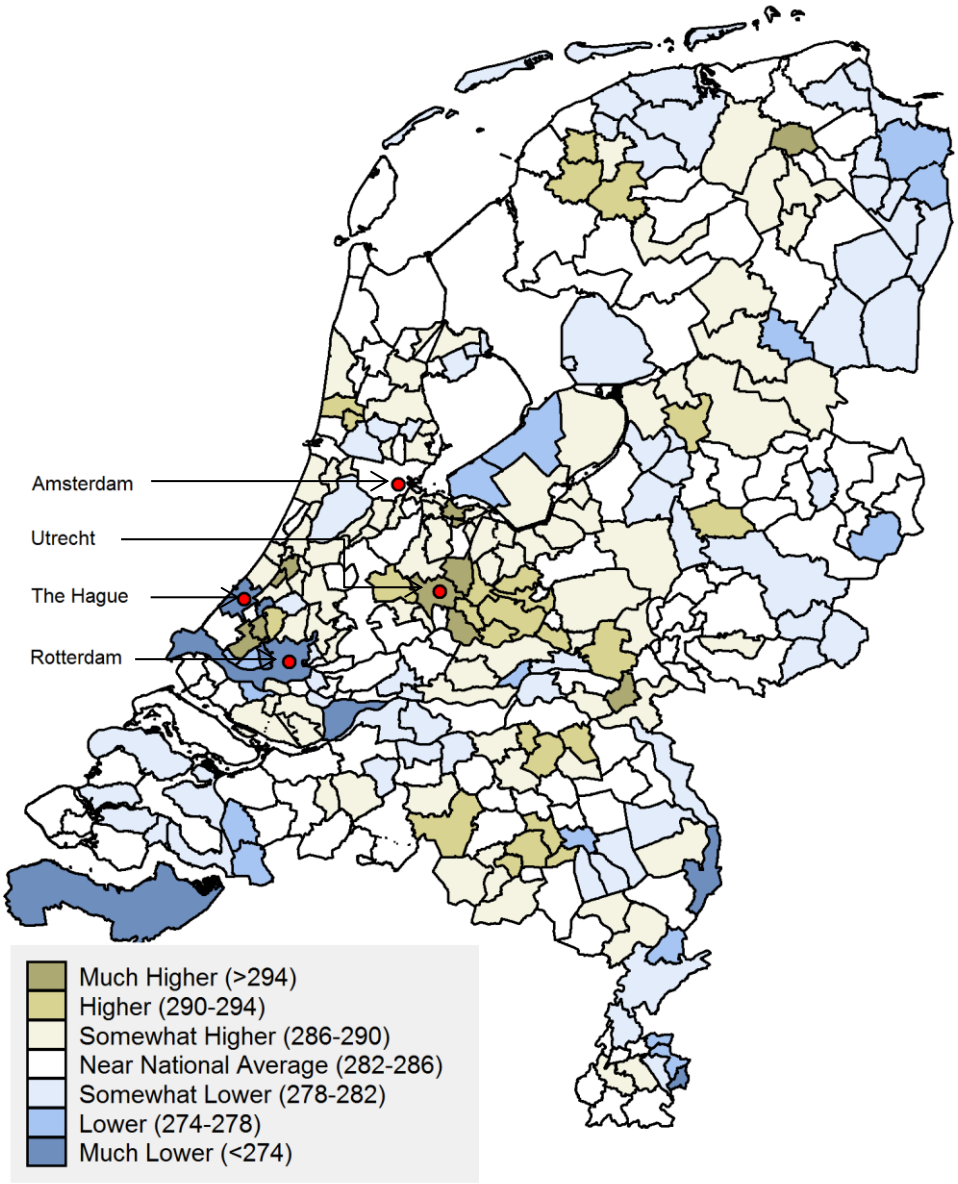


Figure 5: Estimated average literacy scores per municipality

Figure 5 shows the average literacy scores per municipality cluster. Even between close geographical areas, there are often significant differences, highlighting how skills can vary at the local level. Generally, the highest scores for literacy can be found in the center of the country, around the city of Utrecht. Large university cities also do well (Rotterdam being a notable exception). Aside from known problem areas in the western part of the Netherlands, the scores for literacy are low in the peripheral regions.

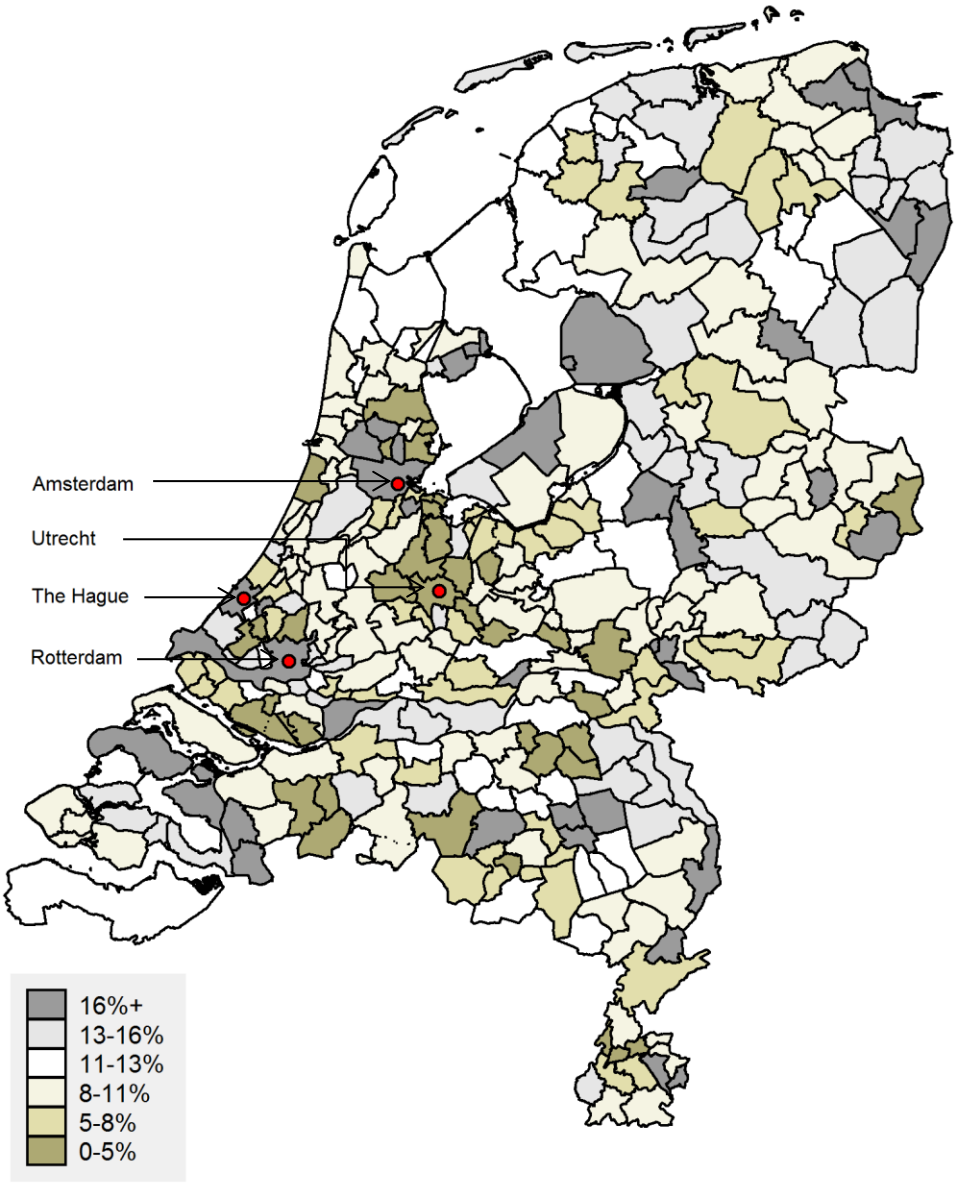


Figure 6: Estimated percentage of individuals classified as having low literacy proficiency scores per municipality

Figure 6 shows the regional results for the percentage of low literates. There is a similar pattern when we look at areas in terms of the percentage of low literates. The first big notable difference, however, is that, in most cases, large cities do much worse in terms of their percentage of low literates in their population, which underlines the usefulness of having both indicators. Low literacy is mainly found in populations with certain characteristics. The average literacy score could give an idea of the overall situation for a population but not how it is distributed. Both measures together provide a more complete picture of the literacy within each area.

Below, we give some examples of how SAE estimates for literacy can relate to other outcomes at the regional level. As stated earlier, low literacy is generally associated with economic disadvantages (such as unemployment and low earnings), but also with negative outcomes in the non-economic domain, such as civic participation, health, and crime. Low literacy can affect such outcomes both directly, through the hampered ability of individuals to make beneficial life choices, and indirectly,

through peer effects in the local neighborhood. Either way, knowledge of regional differences can be a powerful tool for policy interventions aimed at tackling these problems. This is not simply a matter of identifying areas of low literacy, since this is unlikely to be the sole cause of such problems. Policy makers and professionals responsible for policy implementation have an interest in distinguishing regions in which crime, poor health, and other unwanted outcomes are associated with low literacy from regions in which these problems are driven more by other factors. Such knowledge can greatly improve the cost effectiveness of interventions.

As a simple illustration, we plot the relation between (low) literacy and two unwanted non-economic problems: violent crime and obesity. Note that the following is for illustration purposes only. We recognize that real policy interventions will usually be based on more detailed knowledge of the nature of the problem at hand, which could mean that the information presented here is used somewhat differently than proposed here, in combination with other sources of information. By plotting these outcomes with both the direct estimates and the SAE results, we demonstrate how SAE predictions improve relations with other variables by partially eliminating sampling error. This approach facilitates the implementation of more targeted policy interventions.

Figure 7 shows the relation at the regional level between literacy and the incidence of violent crime. Preliminary analysis reveals that the relation is stronger in terms of the proportion of people with low literacy proficiency scores than in terms of average literacy scores, suggesting that the problem may be strongly concentrated at the lower end of the literacy skill continuum.

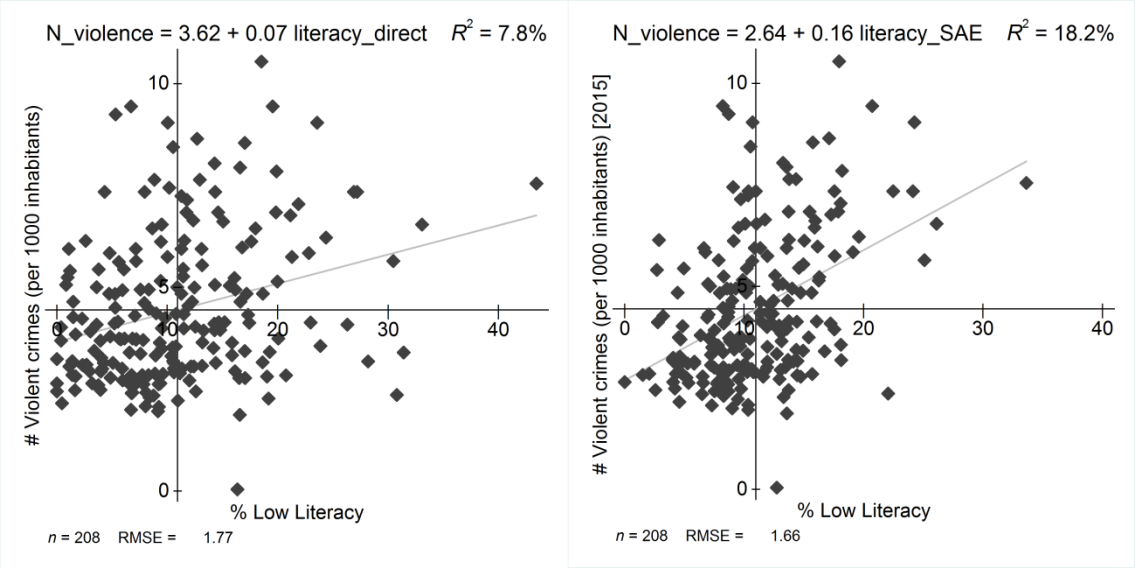


Figure 7: Linear model of the number of violent crimes in 2015 per 1,000 persons versus the percentage of low literates in that region, comparing the results of the direct estimator (left) versus the SAE results (right).

As we can see from Figure 7, the analysis based on direct estimates shows only a weak relation, with an *R-squared* value of only 8%, while the model based on SAE has a much clearer relation, with an *R-squared* value of 18%. This result shows that, at the municipality level, the two variables have a much clearer positive association than one would suspect based on direct observations alone. However, even based on SAE, the relation is far from perfect. The policy response in any given municipality is likely to depend on the position it occupies in the graph. In the lower left quadrant, we see the regions in which both the proportion of inhabitants with low literacy scores and the

incidence of violent crime are low. These are areas in which there is little or no need for intervention. In the lower right quadrant are regions in which low literacy is prevalent but not related with a higher incidence of violent crime. Once again, there is little or no need for policy intervention in these regions, at least in terms of tackling crime rates. In the upper left quadrant are regions with a high rate of violent crime but that are not characterized by strong concentrations of low literates. In these areas there is a need for policy intervention, but little indication that the problem can be tackled by targeting literacy deficits. Only in the upper right quadrant, where a high incidence of violent crime is associated with high concentrations of low literates, do literacy levels loom large as a potentially effective policy lever.

Figure 8 shows the relation between literacy and another important social problem, obesity. In this case, it turns out that the relation at the regional level is stronger in terms of average literacy scores than in terms of the prevalence of low literates. This is already a preliminary clue that the problem could be tackled by investing in improved literacy at all levels.

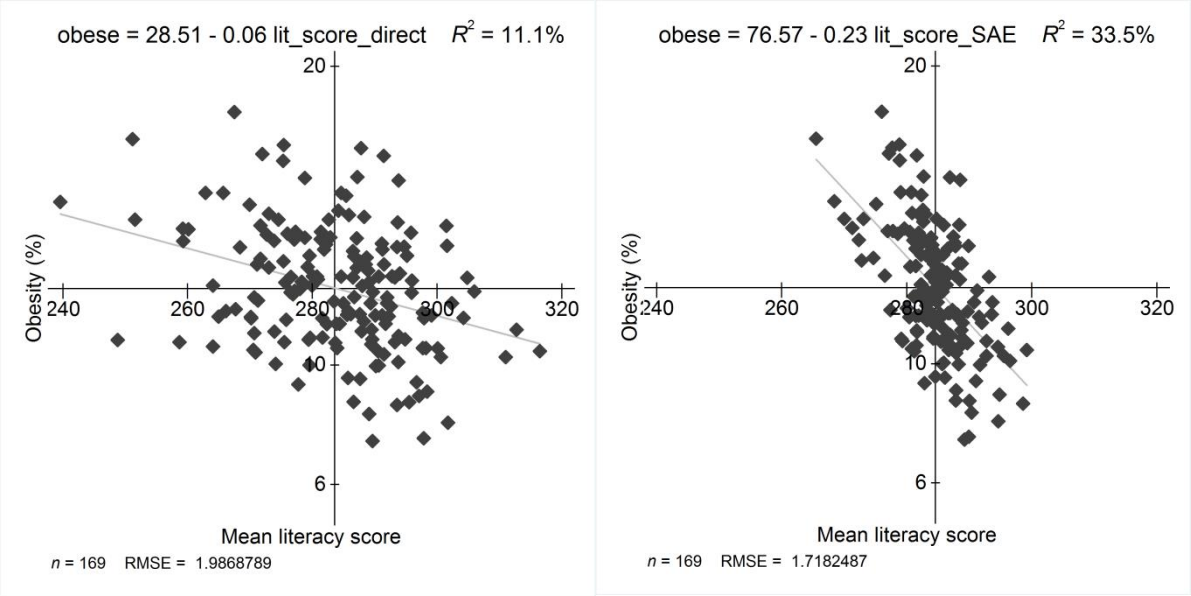


Figure 8: Linear model of the proportion of obese people (in 2012) versus the average literacy level in that region, comparing the results of the direct estimator (left) versus the SAE results (right).

Once again we see that the analysis based on SAE (R^2 = 33.5%) shows a much stronger relation than that based on direct estimates (R^2 = 11.1%), implying that, at the municipality level, the two variables have a much clearer positive association than one would suspect based on direct observations alone. In terms of policy interventions, the position of a given municipality in the graph is indicative of the kind of policy response that could be considered appropriate. There is little incentive to launch literacy-based interventions in the regions in the lower right quadrant, since these are regions with high literacy and a low incidence of obesity. In the lower left and upper right quadrants, literacy-based interventions also do not look promising, since literacy and obesity do not coincide in these regions. Only in the upper left quadrant do we see a high incidence of obesity together with a low average level of literacy. This finding suggests that literacy could potentially be targeted as a policy lever to tackle the problem of obesity in these regions.

7. Conclusion

In this paper, we have combined PIAAC survey data with LFS data to obtain estimates of the literacy levels for municipalities in the Netherlands, both the average literacy scores and the percentage of people with low literacy scores. This was done using SAE models fitted with an HB approach.

Direct estimators only use observations obtained in each specific area to estimate literacy for the area. Results obtained with direct estimators at the regional level, therefore, suffer from small samples sizes for most areas, leading to high standard errors. In this paper, we applied model-based estimation procedures to improve the effective sample size in the different areas, resulting in a considerable improvement of the precision of the estimates of literacy levels, even in larger cities of the Netherlands.

We show that we can obtain reliable estimates at a very detailed regional level by using these SAE techniques. This is important, since policy aimed at reducing low literacy often operates at low regional levels, such as municipalities. We show that we can obtain reliable estimates for the average literacy level and the percentage of low literates for over 200 municipalities in the Netherlands. The findings show that average literacy levels are higher in big cities than in more rural areas, a finding that is consistent with the literature (e.g., McHenry, 2014). However, we also show that large cities cope with higher proportions of low literates, indicating the importance of looking at both measures of literacy.

The estimates can be used for a more optimal allocation of resources to combat low literacy. We also illustrated that more precise SAE estimates are helpful in establishing relations with other variables more clearly. This approach can be used, for example, to identify municipalities that suffer from multiple problems, such as low literacy and health problems or other social problems. In some municipalities, these problems coincide and in some they do not. Identifying the typical mix of problems a municipality is confronted with is key to the development of a successful intervention strategy. The regional estimates for literacy therefore give room for policy makers to implement more directed policies at a detailed regional level.

Future research will focus on the estimation of other skills measured in PIAAC, such as numeracy or problem solving, or by estimating literacy levels in other classifications, such as detailed levels of occupation (for an example, see van der Velden and Bijlsma, 2017). By making these kinds of estimates possible, detailed data become available in areas previously inaccessible due to time and budget constraints.

However, there are a number of caveats to keep in mind when interpreting the results. First and foremost, it must be stressed that these methods rely on statistical model assumptions. Careful model selection and evaluation are therefore an important and necessary part of SAE. The method assumes that the effects of covariates at the regional level are the same as at the national level, with random effects capturing regional differences. While this should hold in most cases, exceptions can occur, such as municipalities with higher returns for schooling or foreign students scoring better than average immigrants. The results should always be viewed with possible local anomalies in mind.

A number of improvements can be made in the estimation of the model. Currently, data used from the LFS are assumed to be the true population means and the corresponding sampling errors are assumed to be negligible. We could use the method of Ybarra and Lohr (2008) to properly take this error term into account. For the percentage of low literates model, a logarithmic model could lead to better estimations between the 0% and 5%, which currently show some bias toward the bottom end of the distribution.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988), County crop areas using survey data and satellite data, *Journal of the American Statistical Association 83* (401), 28-36.

Boonstra, H. J. (2015), *Package 'hbsae'.* Version 1.0, available at https://cran.r-project.org/web/packages/hbsae/hbsae.pdf [9-12-2015].

Buisman, M., Allen, J., Fouarge, D., Houtkoop W. and Van der Velden, R. (2013), *PIAAC: Kernvaardigheden voor werk en leven. Resultaten van de Nederlandse survey 2012*, Den Bosch/Maastricht: ECBO/ROA.

Coulombe, S. and Tremblay, J.F. (2007), Skills, Education, and Canadian Provincial Disparity, *Journal of Regional Science, 47 (5)*, 965–991.

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003), Micro estimation of poverty and inequality, *Econometrica 71*, 355-364.

Fay, R. E. and Herriot, R. A. (1979), Estimates of income for small places: An application of James-Stein procedures to census data, *Journal of the American Statistical Association 74 (366)*, 269-277.

Gibson, A. and Hewson, P. (2012), *2011 Skills for Life Survey: Small Area Estimation Technical Report*. Ref: BIS/12/1318.

Hanushek, E.A. and Woessmann, L. (2008), The Role of Cognitive Skills in Economic Development, *Journal of Economic Literature, 46 (3)*, 607–668.

Hanushek, E.A. and Woessmann, L. (2011), *The Economics of International Differences in Educational Achievement.* Handbook of the Economics of Education, Vol. 3, Amsterdam: North Holland, 89-200.

Johnson, E. G. and Rust, K. F. (1992), Sampling and Weighting in the National Assessment, *Journal of Educational and Behavioral Statistics, 17 (2)*, 111-129.

McHenry, P. (2014), The Geographic Distribution of Human Capital: Measurement of Contributing Mechanisms, *Journal of Regional Science, 54 (2)*, 215–248.

Moretti, E. (2004), Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data, *Journal of Econometrics, 121*, 175–212.

National Research Councel (2000*), Small Area Estimates of School-Age Children in Poverty: Evaluation of current methodology*. C.F. Citro and G. Kalton (eds.). Committee on National Statisitcs, Washington, DC, National Academy Press.

Organisation for Economic Co-operation and Development (2013a), *OECD skills outlook 2013: first results from the survey of adult skills*. OECD Publishing, Paris.

Organisation for Economic Co-operation and Development (2013b), *Technical Report of the Survey of Adult Skills (PIAAC).* Available from http://www.oecd.org/site/piaac/publications.htm [9-12-2015]

Organisation for Economic Co-operation and Development (2013c), *The Survey of Adult Skills - Reader's Companion*. OECD Publishing, Paris.

Potropek, A. and Jabubowski, M. (2013), *Package 'PIAAC tools'.* Available at https://ideas.repec.org/c/boc/bocode/s457728.html [20-09-2016].

Pfeffermann, D. (2013), New Important Developments in Small Area Estimation, *Statistical Science, 28 (1)*, 40–68.

PricewaterhouseCoopers (2013)*, Laaggeletterdheid in Nederland kent aanzienlijke maatschappelijke kosten*. PWC, interne rapportage, Amsterdam.

Rao, J.N.K and Molina, I. (2015), *Small Area Estimation, Second Edition*. John Wiley and Sons, New York.

Rubin, D. B. (1996), Multiple Imputation After 18+ Years, *Journal of the American Statistical Association, 91 (434)*, 473-489.

Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer verlag, New York.

Statistics Netherlands (2014), *Methoden en definities Enquête Beroepsbevolking 2014*. Available at http://www.cbs.nl/NR/rdonlyres/B53DC0DE-743C-4FA2-B7D0-73535C758FEB/0/EBBMethodenendefinities2014.pdf [9-12-2015].

Van den Brakel, J.A. and Krieg, S. (2015), Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design, *Survey Methodology, 41(2)*, 267-296.

Van der Velden, R. and Bijlsma, I. (2017), Skill Effort: A New Theoretical Perspective on the Relation Between Skills, Skill Use, Mismatches, and Wages, ROA Research Memorandum 2017/5, Maastricht: Research Centre for Education and the Labour Market.

Yamamoto, K. (2014), *Using PIAAC Data for Producing Regional Estimates*, working paper, Educational Testing Service, Princeton.

Ybarra, L. M. R. and Lohr, S. L. (2008), Small area estimation when auxiliary information is measured with error, *Biometrika 4 (95)*, 919–931.

Appendix

Table A1: Comparison of weighted dataset averages

Covariate	PIAAC average[1]		LFS average	
Age	41.0	(14.2)	40.6	(14.1)
Male	50%	(50.0)	50%	(50.0)
ISEI08-score	48.7	(18.4)	46.5	(10.6)
Immigrant status				
1st gen	12%	(32.6)	14%	(34.7)
2nd gen	3%	(16.8)	9%	(29.2)
Employment status				
Student	14%	(34.4)	13%	(33.8)
Self-employed	10%	(29.9)	10%	(29.8)
Fulltime employee	37%	(48.3)	31%	(46.2)
Parttime employee	20%	(40.3)	22%	(41.2)
Other	3%	(17.1)	4%	(19.0)
Education[2]				
Vocational ed.	57%	(49.4)	58%	(49.4)
Years of schooling	13.2	(3.7)	13.4	(3.6)

[1] For the Netherlands, the control variables that were used to calibrate weights in PIAAC are: Gender by age (10), origin by generation (5), group of provinces by degree of urbanization (18), household type (5), social status by income (25), term of registration in population registry (2), percentage of high level education by percentage of low level education (18).

[2] The education variables contained slightly more than 1% missing values. For area estimates, missing values are assumed to have the same distribution as the known values.

Figure A1: Q-Q Plots of the SAE estimates (left), residuals (middle) and random effects (right); of the unit-level estimates of the estimated literacy scores (top) and the area-level estimates of the percentage of low literates after benchmarking (bottom).

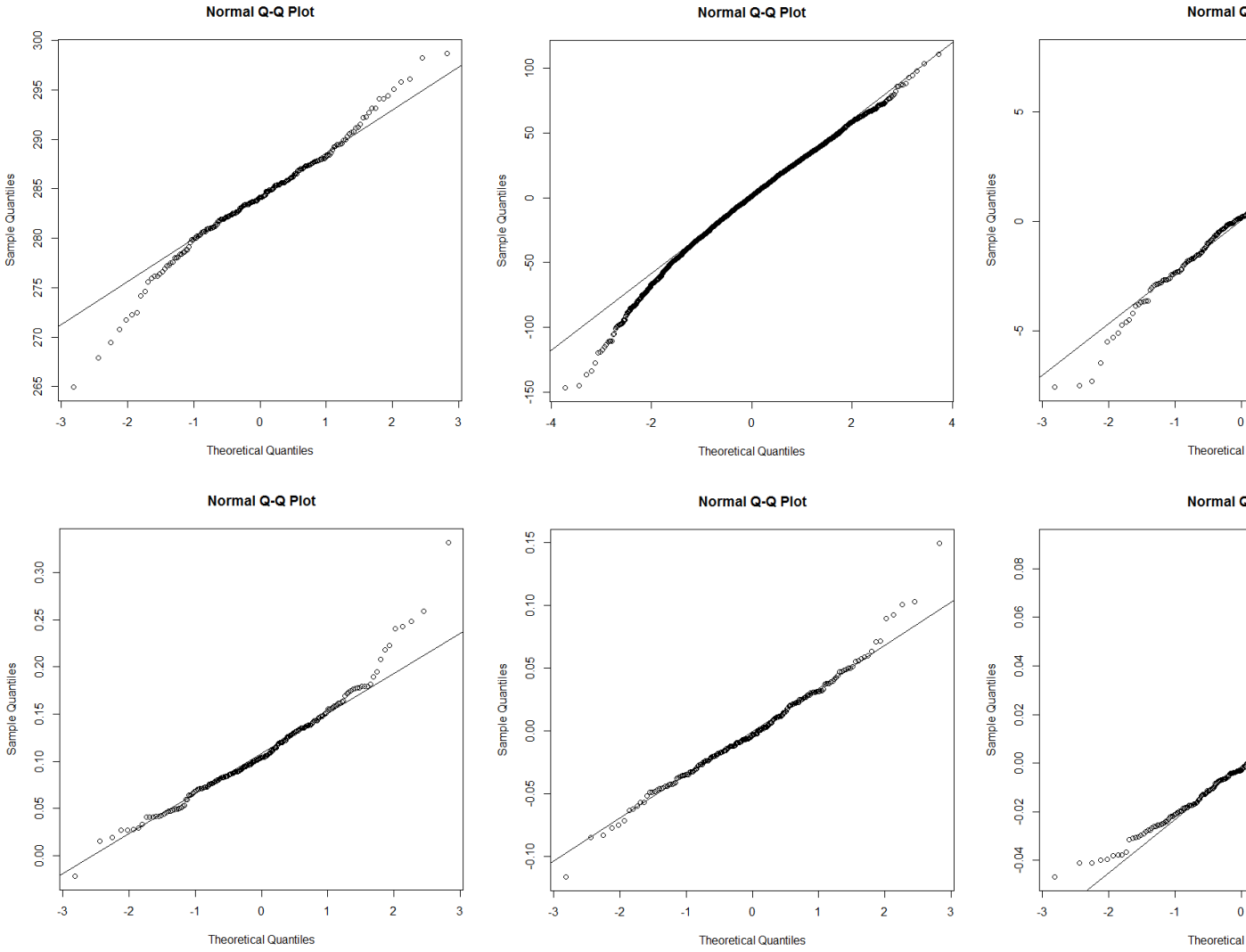


Table A2: Literacy per municipality

Province	Municipality cluster	Average Score (Std. Err)	% Low Literates (Std.Err)
Groningen	Bellingwedde/Oldambt	278.0 (4.0)	15.0% (0,035)
	Menterwolde/Pekela/Veendam	278.5 (3.9)	16.0% (0,035)
	Stadskanaal/Vlagtwedde	278.5 (4.0)	17.3% (0,037)
	Appingedam/Delfzijl/Loppersum	279.3 (3.9)	19.6% (0,033)
	Bedum/De Marne/Eemsmond/Ten Boer/Winsum	285.8 (3.8)	9.4% (0,033)
	Grootegast/Leek/Marum/Zuidhorn	288.0 (3.0)	7.8% (0,033)
	Haren/Hoogezand-Sappemeer/Slochteren	284.6 (3.9)	8.9% (0,032)
	Groningen	298.6 (3.9)	10.5% (0,025)
Friesland	Dongeradeel/Ferwerderadiel/Schiermonnikoog/i.a.	280.0 (4.1)	15.7% (0,036)
	Leeuwarderadeel/Tytsjerksteradiel	281.5 (4.0)	11.6% (0,035)
	Leeuwarden	286.1 (3.7)	15.8% (0,030)
	Achtkarspelen/Dantumadiel/Kollumerland Ca	281.1 (4.3)	13.7% (0,033)

	Franekeradeel/Harlingen/Het Bildt	283.2 (3.6)	12.7% (0,033)
	Boarnsterhim/Littenseradiel/Menameradiel	291.0 (4.0)	7.3% (0,036)
	Gaasterlan-Sleat/Lemsterland/Sudwest Fryslan	285.7 (3.9)	12.9% (0,029)
	Heerenveen	288.4 (3.8)	14.0% (0,033)
	Smallingerland	282.4 (4.1)	17.6% (0,034)
	Skarsterlan/Weststellingwerf	285.4 (3.9)	10.3% (0,035)
	Ooststellingwerf/Opsterland	285.2 (3.8)	13.5% (0,032)
Drenthe	Aa en Hunze/Midden-Drenthe	285.4 (3.3)	12.7% (0,031)
	Noordenveld/Tynaarlo	288.2 (3.5)	7.3% (0,033)
	Assen	286.1 (3.8)	11.7% (0,033)
	Borger-Odoorn/Coevorden	281.1 (3.9)	14.4% (0,033)
	Emmen	278.0 (3.8)	13.7% (0,028)
	Hoogeveen	276.7 (3.8)	16.3% (0,033)
	De Wolden/Meppel/Westerveld	287.3 (3.9)	9.5% (0,031)
Overijssel	Zwolle	291.2 (3.8)	10.1% (0,027)
	Hardenberg	288.1 (3.8)	9.2% (0,031)
	Kampen	287.5 (3.6)	13.1% (0,039)
	Steenwijkerland	283.9 (3.8)	15.7% (0,035)
	Dalfsen/Ommen/Staphorst/Zwartewaterland	287.5 (3.5)	6.5% (0,028)
	Olst-Wijhe/Raalte	283.3 (4.0)	13.6% (0,037)
	Deventer	290.1 (3.8)	6.8% (0,032)
	Almelo	280.4 (4.0)	19.1% (0,032)
	Hengelo	284.0 (4.2)	7.9% (0,029)
	Enschede/Scherpenzeel	276.7 (3.5)	25.1% (0,027)
	Borne/Dinkelland/Tubbergen	283.6 (3.6)	9.5% (0,030)
	Losser/Oldenzaal	284.1 (4.0)	4.6% (0,034)
	Hellendoorn/Twenterand	282.3 (3.5)	8.7% (0,031)
	Haaksbergen/Hof van Twente	285.2 (3.9)	8.0% (0,032)
	Rijssen-Holten/Wierden	283.6 (3.8)	10.3% (0,032)
Gelderland	Ede/Wageningen	288.9 (3.7)	8.7% (0,029)
	Apeldoorn	286.6 (3.6)	12.2% (0,026)
	Barneveld	284.6 (3.9)	12.3% (0,034)
	Harderwijk	287.3 (4.2)	9.0% (0,038)
	Nijkerk	287.5 (4.0)	7.8% (0,038)
	Epe/Voorst	281.1 (3.9)	18.1% (0,034)
	Hattem/Heerde/Oldebroek	281.9 (3.3)	14.7% (0,037)
	Elburg/Nunspeet	283.9 (3.8)	12.0% (0,034)
	Ermelo/Putten	288.3 (3.8)	7.0% (0,033)
	Berkelland/Lochem	281.9 (3.9)	14.4% (0,032)
	Zutphen	284.9 (3.3)	15.0% (0,033)
	Doetinchem	287.7 (4.1)	6.6% (0,033)
	Bronckhorst/Brummen	284.1 (3.9)	10.3% (0,032)
	Montferland/Oude IJsselstreek	285.9 (3.4)	5.8% (0,031)
	Aalten/Oost Gelre/Winterswijk	280.8 (3.6)	14.4% (0,029)
	Arnhem	286.3 (3.8)	8.7% (0,027)
	Nijmegen	295.5 (3.8)	10.4% (0,026)
	Wijchen	288.8 (3.9)	4.5% (0,036)
	Rheden/Rozendaal	283.8 (4.0)	8.3% (0,040)
	Lingewaard	285.9 (3.9)	5.9% (0,036)
	Overbetuwe/Renkum	290.4 (4.0)	2.8% (0,029)
	Doesburg/Zevenaar	283.0 (4.1)	8.3% (0,034)

	Groesbeek/Heumen/Millingen aan de Rijn/Ubbergen	287.3 (4.0)	7.0% (0,036)
	Beuningen/Druten	284.1 (4.1)	11.7% (0,036)
	Duiven/Rijnwaarden/Westervoort	282.4 (4.0)	17.2% (0,033)
	Tiel	276.5 (3.5)	17.6% (0,037)
	Buren/Culemborg	286.3 (3.9)	12.4% (0,035)
	Maasdriel/Zaltbommel	283.8 (4.0)	13.2% (0,034)
	Geldermalsen/Lingewaal/Neerijnen	286.0 (4.0)	6.9% (0,034)
	Neder-Betuwe/West Maas en Waal	281.6 (4.1)	9.7% (0,035)
Utrecht	De Bilt	294.6 (3.9)	4.1% (0,040)
	De Ronde Venen	285.6 (4.1)	8.2% (0,036)
	Soest	287.9 (4.0)	4.1% (0,037)
	Utrechtse Heuvelrug	290.8 (3.9)	8.5% (0,036)
	Houten	294.8 (4.1)	6.5% (0,035)
	Woerden	292.8 (4.1)	2.6% (0,037)
	Nieuwegein	284.8 (4.0)	13.8% (0,031)
	Zeist	292.7 (4.1)	10.9% (0,039)
	Veenendaal	283.3 (3.9)	8.7% (0,034)
	Stichtse Vecht	288.1 (4.2)	4.1% (0,034)
	Amersfoort	286.7 (4.2)	8.6% (0,025)
	Utrecht	299.2 (4.0)	2.9% (0,021)
	Baarn/Bunschoten/Eemnes	288.7 (3.7)	7.2% (0,039)
	Leusden/Renswoude/Woudenberg	291.7 (3.9)	9.0% (0,034)
	Bunnik/Rhenen/Wijk bij Duurstede	291.6 (4.1)	4.6% (0,034)
	Ijsselstein/Montfoort	288.2 (2.8)	7.5% (0,037)
	Lopik/Oudewater/Vianen	284.3 (4.0)	8.3% (0,036)
Noord-Holland	Koggenland/Medemblik	286.1 (3.3)	10.5% (0,033)
	Hollands Kroon/Opmeer/Texel	284.7 (3.9)	11.9% (0,032)
	Schagen	282.8 (3.9)	12.9% (0,035)
	Den Helder	282.2 (3.7)	10.7% (0,034)
	Hoorn	284.2 (3.7)	14.4% (0,031)
	Drechterland/Enkhuizen/Stede Broec	281.6 (3.4)	17.6% (0,033)
	Heerhugowaard/Langedijk/Schermer	285.7 (3.7)	9.1% (0,031)
	Alkmaar	282.9 (3.7)	11.0% (0,027)
	Bergen NH/Heiloo	289.9 (3.8)	9.7% (0,037)
	Velsen	285.8 (4.0)	10.2% (0,031)
	Beverwijk/Heemskerk	283.8 (4.0)	8.7% (0,030)
	Castricum/Uitgeest	290.0 (4.0)	8.3% (0,036)
	Haarlem/Haarlemmerliede Ca	289.0 (3.6)	9.7% (0,026)
	Bloemendaal/Heemstede/Zandvoort	289.2 (3.3)	3.2% (0,038)
	Wormerland/Zaanstad	278.9 (3.6)	16.3% (0,025)
	Purmerend	281.5 (3.8)	9.5% (0,030)
	Amstelveen/Diemen/Ouder-Amstel	289.4 (4.0)	8.0% (0,030)
	Haarlemmermeer	280.4 (3.5)	13.7% (0,028)
	Amsterdam	284.3 (3.3)	18.0% (0,018)
	Landsmeer/Beemster/Edam-Volendam/i.a.	287.9 (2.4)	4.0% (0,032)
	Aalsmeer/Uithoorn	288.4 (4.1)	8.3% (0,037)
	Blaricum/Huizen/Laren	287.9 (2.7)	6.4% (0,035)
	Hilversum	283.9 (3.6)	14.7% (0,031)
	Muiden/Weesp/Wijdemeren	286.9 (3.6)	1.5% (0,039)
	Bussum/Naarden	294.6 (3.9)	0.0% (0,040)
Zuid-Holland	Katwijk/Oegstgeest	287.4 (4.2)	13.1% (0,033)

	Leiden/Voorschoten	296.5 (3.2)	10.5% (0,026)
	Hillegom/Lisse/Teylingen	289.7 (4.2)	10.1% (0,030)
	Kaag en Braassem/Leiderdorp/Zoeterwoude	287.0 (3.8)	8.7% (0,033)
	Noordwijk/Noordwijkerhout	284.2 (3.7)	10.3% (0,036)
	Zoetermeer	281.4 (3.9)	13.6% (0,029)
	s Gravenhage	272.3 (4.1)	24.3% (0,020)
	Pijnacker-Nootdorp	290.4 (4.0)	7.8% (0,035)
	Leidschendam-Voorburg/Wassenaar	289.7 (3.7)	7.1% (0,032)
	Rijswijk	282.3 (3.7)	12.0% (0,042)
	Delft/Midden-Delfland	296.2 (3.3)	2.7% (0,029)
	Westland	282.2 (3.4)	13.1% (0,028)
	Gouda	282.6 (4.0)	11.0% (0,030)
	Alphen aan den Rijn	284.5 (3.9)	12.0% (0,031)
	Nieuwkoop/Schoonhoven/Vlist/i.a.	285.3 (4.3)	10.6% (0,027)
	Boskoop/Rijnwoude/Waddinxveen	286.6 (3.4)	8.8% (0,032)
	Ouderkerk/Zuidplas	288.5 (3.6)	9.8% (0,036)
	Ridderkerk	278.8 (4.2)	11.9% (0,037)
	Barendrecht	283.1 (3.5)	14.1% (0,041)
	Lansingerland	287.9 (3.9)	2.1% (0,037)
	Capelle aan den IJssel	279.3 (4.1)	15.9% (0,037)
	Maassluis/Vlaardingen	274.6 (4.1)	13.7% (0,027)
	Spijkenisse	276.0 (3.7)	12.7% (0,031)
	Schiedam	268.4 (3.8)	33.6% (0,030)
	Rotterdam	273.1 (4.1)	20.7% (0,019)
	Goeree-Overflakkee	285.4 (3.6)	10.0% (0,037)
	Bernisse/Brielle/Hellevoetsluis/Westvoorne	283.9 (3.9)	7.1% (0,030)
	Cromstrijen/Korendijk/Oud-Beijerland/Strijen	288.5 (2.6)	4.2% (0,033)
	Krimpen aan den IJssel/Nederlek	283.2 (3.7)	13.7% (0,037)
	Albrandswaard/Binnenmaas	286.9 (4.1)	8.3% (0,033)
	Zwijndrecht	279.0 (4.0)	15.4% (0,036)
	Dordrecht	271.3 (4.0)	22.5% (0,029)
	Giessenlanden/Gorinchem	280.5 (3.7)	11.4% (0,036)
	Alblasserdam/Hendrik-Ido-Ambacht	285.2 (3.8)	5.2% (0,035)
	Hardinxveld-Giessendam/Papendrecht/Sliedrecht	282.6 (3.7)	7.5% (0,031)
	Leerdam/Molenwaard/Zederik	282.4 (4.2)	8.7% (0,031)
Zeeland	Hulst/Sluis/Terneuzen	272.8 (3.8)	12.6% (0,030)
	Borsele/Vlissingen	283.0 (3.6)	10.7% (0,035)
	Middelburg/Veere	282.9 (3.7)	9.9% (0,032)
	Goes/Kapelle/Noord Beveland/Reimerswaal	281.4 (3.9)	14.6% (0,031)
	Schouwen-Duiveland/Tholen	280.7 (4.2)	16.7% (0,037)
Noord-Brabant	Etten-Leur	286.4 (3.7)	4.4% (0,038)
	Oosterhout	280.5 (4.1)	10.2% (0,038)
	Bergen op Zoom/Woensdrecht	277.4 (3.8)	16.1% (0,030)
	Roosendaal	282.4 (3.2)	10.4% (0,033)
	Breda	286.1 (4.2)	13.6% (0,025)
	Halderberge/Rucphen/Zundert	282.6 (4.0)	4.9% (0,033)
	Moerdijk/Steenbergen	283.9 (3.0)	8.7% (0,036)
	Drimmelen/Geertruidenberg	282.7 (4.2)	7.2% (0,037)
	Waalwijk	280.7 (3.9)	11.4% (0,033)
	Tilburg	286.9 (3.9)	13.3% (0,022)
	Goirle/Hilvarenbeek/Oisterwijk	293.1 (4.2)	4.7% (0,032)

	Aalburg/Werkendam/Woudrichem	279.6 (3.9)	13.3% (0,034)
	Alphen-Chaam/Baarle-Nassau/Gilze en Rijen	284.7 (4.0)	8.8% (0,037)
	Dongen/Loon op Zand	283.0 (4.1)	7.5% (0,035)
	Uden	286.6 (3.9)	5.0% (0,035)
	Heusden	281.6 (4.0)	9.6% (0,036)
	Oss	284.8 (4.1)	12.5% (0,029)
	s Hertogenbosch	287.8 (3.8)	9.1% (0,026)
	Boxtel/Haaren/Vught	285.8 (3.3)	11.3% (0,031)
	Sint-Oedenrode/Veghel	285.3 (3.7)	13.0% (0,037)
	Boekel/Boxmeer/Sint Anthonis	285.9 (3.7)	14.5% (0,034)
	Cuijk/Grave/Mill en Sint Hubert	283.0 (3.9)	14.9% (0,033)
	Bernheze/Landerd/Maasdonk	291.2 (4.0)	4.8% (0,032)
	Schijndel/Sint-Michielsgestel	286.5 (4.0)	8.1% (0,035)
	Veldhoven	293.6 (4.0)	4.3% (0,035)
	Helmond	277.8 (3.8)	17.1% (0,033)
	Eindhoven	293.7 (3.9)	8.2% (0,023)
	Gemert-Bakel/Laarbeek	282.4 (3.7)	18.1% (0,035)
	Geldrop-Mierlo/Nuenen Ca/Son en Breugel	292.0 (3.1)	7.5% (0,031)
	Asten/Deurne/Someren	281.2 (3.8)	12.9% (0,034)
	Cranendonck/Heeze-Leende/Waalre	284.3 (4.1)	6.6% (0,033)
	Bladel/Eersel/Reusel-De Mierden	289.9 (3.9)	7.2% (0,034)
	Best/Oirschot	284.3 (3.9)	16.1% (0,032)
	Bergeijk/Valkenswaard	287.8 (4.1)	11.9% (0,033)
Limburg	Horst aan de Maas	288.4 (4.0)	8.4% (0,037)
	Bergen LB/Gennep/Mook en Middelaar/Venray	281.5 (3.6)	13.6% (0,030)
	Beesel/Peel en Maas	284.6 (3.6)	10.3% (0,032)
	Venlo	270.0 (3.8)	24.1% (0,028)
	Weert	283.5 (4.1)	11.3% (0,035)
	Roermond	275.1 (3.8)	18.1% (0,034)
	Leudal/Nederweert	286.3 (3.9)	10.0% (0,034)
	Echt-Susteren/Maasgouw/Roerdalen	281.5 (4.1)	5.3% (0,033)
	Kerkrade	265.5 (3.9)	26.1% (0,035)
	Heerlen	278.6 (3.8)	18.2% (0,032)
	Sittard-Geleen	280.8 (3.6)	9.7% (0,030)
	Maastricht	284.6 (3.3)	13.8% (0,027)
	Beek/Schinnen/Stein	284.1 (3.9)	4.7% (0,032)
	Brunssum/Landgraaf/Onderbanken	277.7 (3.8)	9.3% (0,034)
	Eijsden-Margraten/Gulpen-Wittem/Simpelveld/Vaals	285.2 (4.0)	9.5% (0,035)
	Meerssen/Nuth/Valkenburg aan de Geul/Voerendaal	286.1 (3.9)	5.3% (0,034)
Flevoland	Dronten/Zeewolde	288.3 (3.8)	10.8% (0,033)
	Noordoostpolder/Urk	279.1 (4.1)	22.1% (0,034)
	Lelystad	277.1 (3.8)	17.9% (0,031)
	Almere	276.9 (3.3)	15.5% (0,026)