Small Area Proficiency Estimation
Extension of the PIAAC Population Modeling for Proficiency Estimation without Direct Assessment
Kentaro Yamamoto, ETS
Outline of Presentation

• What is Small Area Estimation and Why It is Important?
• Distinguishing Individual Testing and Large Scale Assessments
• Use of Population Modeling in PIAAC
• Extension of Population Modeling to Small Area Estimation
• Future Applications
Why Small Area Estimation (SAE)?

- PIAAC provides extensive data useful for researchers and policy makers on hundreds of cognitive tasks and background questions from large national samples.
- PIAAC requires effort, money and time in order to develop instruments, conduct survey and reporting making it difficult to conduct frequently.

- Policy makers are interested: to know the proficiency distribution of those who did not participate in PIAAC to know the proficiency distribution of some population in the future.

- SAE methodology provides policy makers with information they seek while reducing burden on time, money and effort.
Previous Efforts on Small Area Estimation

- Mislevy, R. J. (1985)
- Reder, S. (1994)
### Moving from Individual Testing to Small Area Estimation

<table>
<thead>
<tr>
<th>Individual Testing</th>
<th>Population Assessment</th>
<th>Small Area Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Skill accuracy for each respondent</td>
<td>• Accuracy of skill distributions of groups of respondents</td>
<td>• Accuracy of skill distributions of groups of respondents</td>
</tr>
<tr>
<td>• Large Cognitive Data</td>
<td>• Responses are on a sample of items drawn from a pool of items representing construct</td>
<td>• Population Model parameters Statistics of population characteristics</td>
</tr>
<tr>
<td></td>
<td>• Responses to extensive background questions</td>
<td></td>
</tr>
</tbody>
</table>

**Individual Testing**
- Skill accuracy for each respondent
- Large Cognitive Data

**Population Assessment**
- Accuracy of skill distributions of groups of respondents
- Responses are on a sample of items drawn from a pool of items representing construct
- Responses to extensive background questions

**Small Area Estimation**
- Accuracy of skill distributions of groups of respondents
- Population Model parameters Statistics of population characteristics

*ETS* - Listening. Learning. Leading.*

*OECD* - Better Policies for Better Lives
Accurate Estimation of Individual Skill

- Accurately evaluating the proficiency of an individual for diagnosis, selection or placement by administering many cognitive items that represent the domain of testing as completely as possible.

\[ P(x|\theta,\beta,\alpha) \]
Population Assessment is to Estimate Distributions of Subpopulations

- Accurately evaluating proficiency distributions of sampled respondents to provide empirically grounded interpretations to inform policy decisions by administering sampled cognitive items and background questions concerning skill development, social outcomes, and demographic characteristics.

![Diagram showing the relationship between sampled national population, responses to cognitive items, and subpopulation skill distribution.](image)

\[ P(\chi|\theta, \beta, \alpha) \]
Reducing Cognitive Items Increases Measurement Errors

- Sampled items instead of all items that represent assessment construct are administered to each respondent.

- Reduced information results in increased measurement errors.

Sampled National population

Reduced Responses to cognitive items

Subpopulation skill distribution

\[ P(x|\theta, \beta, \alpha) \]
Population Modeling:
Incorporation of Background Information and Cognitive Responses to Represent Proficiencies
Incorporating BQ Restores Accuracy of Distribution Estimates

- BQ data augment the reduced to cognitive data to increase the accuracy of estimating the subpopulation distributions.

Sampled National population

Responses to BQ

Reduced responses to cognitive items

Subpopulation skill distribution

\[ P(\theta|x, y, \Gamma, \Sigma) \]

\[ P(x|\theta, \beta, \alpha) \]

\[ N(\gamma \Gamma, \Sigma) \]
PIAAC Analysis is Based on Cognitive Data and BQ

Responses to BQ

Responses to cognitive items

Subpopulation skill distributions

National population
Completed PIAAC Database for Analyses

- Accurately evaluating proficiency distributions of sampled respondents to provide empirically grounded interpretations to inform policy decisions by administering sampled cognitive items and background questions concerning skill development, social outcomes, and demographic characteristics.

Reduced responses to cognitive items Responses to BQ Sampled National population Subpopulation skill distribution
1. Based on the most strongly related BQ variables and item responses produce population model.

Responses to BQ

Reduced responses to cognitive items

Subpopulation skill distribution

\[ P(\theta|x, y, \Gamma, \Sigma) \]

\[ N(\gamma \Gamma, \Sigma) \]

\[ P(x|\theta, \beta, \alpha) \]
Evaluating Small Area Estimation Feasibility

1. Based on the most strongly related BQ variables and item responses produce population model.
2. Eliminate cognitive responses
3. Estimate subpopulation skill distribution

Responses to BQ
No responses to cognitive items
Subpopulation skill distribution

Strongly related to skill

\[
\mathcal{N}(\gamma \Gamma, \Sigma)
\]

\[
P(\theta | y, \Gamma, \Sigma)
\]
Evaluating Small Area Estimation Feasibility

1. Based on the most strongly related BQ variables and item responses produce population model parameters $(\Gamma, \Sigma)$.
   - Incorporate external data on common background variables
   - Still no cognitive responses

Statistics on selected common variables

Population model parameters from survey

No responses to cognitive items

Projected Subpopulation skill distribution

\[ P(\theta | y, \Gamma, \Sigma) \]
Data for Population Assessment: Reality of Data

- Responses to BQ
- Responses to cognitive items
- Skill distribution

Sampled National population

- Less related to skill
- Strongly related to skill

>5 cognitive response

<5 cognitive responses
Canada collected over 27,000 samples. Over 21,000 cases in English alone. The accuracy of small area estimation for the subset of samples under reduced data can be evaluated against actually observed data. Seven background questions were selected for consideration: Gender, Age, Education, Native/non-Native, Employment Status, Province, and PIAAC routing.

<table>
<thead>
<tr>
<th>Model</th>
<th>Background variables</th>
<th>Cognitive items</th>
<th>Population statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Reported Full 1Population Model</td>
<td>All</td>
<td>All</td>
<td>Canada-E</td>
</tr>
<tr>
<td>2) 7BQ/IR</td>
<td>7BQ</td>
<td>All</td>
<td>Canada-E</td>
</tr>
<tr>
<td>3) 7BQ/no IR</td>
<td>7BQ</td>
<td>NO</td>
<td>Canada-E</td>
</tr>
<tr>
<td>4) Full BQ/IR Ontario</td>
<td>All</td>
<td>All</td>
<td>Ontario</td>
</tr>
<tr>
<td>5) 7BQ/IR Ontario</td>
<td>7BQ</td>
<td>All</td>
<td>Ontario</td>
</tr>
<tr>
<td>6) 7BQ/no IR Ontario</td>
<td>7BQ</td>
<td>NO</td>
<td>Ontario</td>
</tr>
</tbody>
</table>
Adequacy of 7BQ for Population Modeling

- 7BQ is adequate for population modeling for a group with >5 cognitive responses.
- 7BQ is not adequate for a group with <5 cognitive responses.
**Evaluation of the Methodology on Canada-E Data**

- 7BQ without item responses reproduced the best estimates with all BQ and all cognitive responses for national Canadian English samples.

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Background variables</th>
<th>Cognitive items</th>
<th>Population statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Reported Full Population Model</td>
<td>All</td>
<td>All</td>
<td>Canada-E</td>
</tr>
<tr>
<td>2) 7BQ/IR</td>
<td>7BQ</td>
<td>All</td>
<td>Canada-E</td>
</tr>
<tr>
<td>3) 7BQ/no IR</td>
<td>7BQ</td>
<td>NO</td>
<td>Canada-E</td>
</tr>
<tr>
<td>4) Full BQ/IR Ontario</td>
<td>All</td>
<td>All</td>
<td>Ontario</td>
</tr>
<tr>
<td>5) 7BQ/IR Ontario</td>
<td>7BQ</td>
<td>All</td>
<td>Ontario</td>
</tr>
<tr>
<td>6) 7BQ/no IR Ontario</td>
<td>7BQ</td>
<td>NO</td>
<td>Ontario</td>
</tr>
</tbody>
</table>
Evaluation of the Methodology on Canada-E Data

- 7BQ without item responses reproduced the best estimates with all BQ and all cognitive responses for national Canadian English samples.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Background Variables</th>
<th>Cognitive Items</th>
<th>Population Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Reported Full Population Model</td>
<td>All</td>
<td>All</td>
<td>Canada-E</td>
</tr>
<tr>
<td>2) 7BQ/IR</td>
<td>7BQ</td>
<td>All</td>
<td>Canada-E</td>
</tr>
<tr>
<td>3) 7BQ/no IR</td>
<td>7BQ</td>
<td>NO</td>
<td>Canada-E</td>
</tr>
<tr>
<td>4) Full BQ/IR Ontario</td>
<td>All</td>
<td>All</td>
<td>Ontario</td>
</tr>
<tr>
<td>5) 7BQ/IR Ontario</td>
<td>7BQ</td>
<td>All</td>
<td>Ontario</td>
</tr>
<tr>
<td>6) 7BQ/no IR Ontario</td>
<td>7BQ</td>
<td>NO</td>
<td>Ontario</td>
</tr>
</tbody>
</table>

Accuracy of Small Area Estimation Procedures Against The Best Canada-English
Evaluation of the Methodology on Ontario-E Data

- 7BQ without item responses reproduced the best estimates with all BQ and all cognitive responses for Ontario even though the population model was based on the national Canadian English data.
Observations of Estimation Errors

- Subpopulation means and cumulative distributions were well estimated for those with greater than 5 cognitive responses based on the population model of 7 BQ and no item responses.

- Larger subgroups means and cumulative distributions for combined populations regardless of amount of cognitive responses were well estimated for both English speaking populations of Canada and Ontario.

- Estimation errors on the means were greater (>10) for some small groups such as those who do not report level of education (1.4%), those with PhD (0.6%), permanently disabled (3.4%), and not so small groups such as immigrant (18.9%), and failed ICT core (6.2%) and refused to take CBA.
Approaches to Increase Accuracy

- Further development on the population model can be realized using the full interaction among background variables to increase accuracy.
- 7 variables are not sufficient for some of the groups to capture covariance with proficiencies for those who respond less than 5 cognitive items. Search for the best combination of BQ variables.
- All groups with large errors have often 50% or more who responded less than 5 cognitive responses, i.e., not sufficient information to estimate population parameters.
Future Applications

• Estimate proficiency distributions of a group of special interest, a region, and a state based on the national population model.

• Smaller population model with reduced BQ enables us to use a subset of national data to develop unique population model for further customization as well as optimization.

• Use of census information to monitor changes in demographic characteristics to be reflected in proficiency distributions based on the optimized population model based on the common variables.
• Large Scale Assessments such as PIAAC provide a reliable and valid database that can be used by researchers and others to inform policy makers. Because they require a lot of resources to develop and implement there is some concern about the resources needed to conduct them frequently. This presentation provides a conceptual approach to Small Area Estimation using population modeling based on direct assessments, and demonstrates that it can provide reliable and valid estimates of proficiency distributions with much less demands on resources.
Thank you for your attention