IEA’s Rosetta Stone: Implementation and Results

9th Meeting of the Global Alliance to Monitor Learning (GAML)
23 November 2022

Oliver Neuschmidt (IEA)
Rosetta is a collaboration between UIS, IEA, the TIMSS & PIRLS ISC at Boston College, the coordination centers for regional studies, and participating countries.

**Objectives:**

- **Establish a concordance** between ERCE & PASEC regional assessment results and the global TIMSS and PIRLS scales
  - by providing confidence intervals for regional assessment results projected onto the TIMSS mathematics and PIRLS reading scales
- Results allow for the estimation of the proportion of primary school students in regional assessments that achieve TIMSS and PIRLS proficiency levels
Implementation (ERCE: 2019 / PASEC: 2020)

• First successful administration in
  – Colombia and Guatemala: Concordance between TIMSS/PIRLS and ERCE
  – Senegal, Guinea, and Burundi: Concordance between TIMSS/PIRLS and PASEC
• Students were administered their regional assessment (PASEC or ERCE) followed by TIMSS and PIRLS linking booklets in a separate session
  – Material was administered to students in Grade 6
  – Linking booklets contained TIMSS and PIRLS item blocks from Grade 4
→ Combined data was used to construct the concordance tables
### Rosetta Stone – samples

<table>
<thead>
<tr>
<th>ERCE Countries</th>
<th>N (students)</th>
<th>N (for analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colombia</td>
<td>4,467</td>
<td>3,108</td>
</tr>
<tr>
<td>Guatemala</td>
<td>4,894</td>
<td>4,716</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>9,361</strong></td>
<td><strong>7,824</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PASEC Countries</th>
<th>N (students)</th>
<th>N (for analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burundi</td>
<td>2,304</td>
<td>2,271</td>
</tr>
<tr>
<td>Guinea</td>
<td>2,252</td>
<td>2,207</td>
</tr>
<tr>
<td>Senegal</td>
<td>2,072</td>
<td>2,059</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>6,628</strong></td>
<td><strong>6,573</strong></td>
</tr>
</tbody>
</table>

- **ERCE**: Reduced sample size in Colombia: 1,357 students did not receive TIMSS and PIRLS linking booklets
  -> No impact of the reduced sample size on results
- **PASEC**: Rosetta Stone took place one year after PASEC 2019 administration with about half sample size
  -> both samples are comparable (similar distribution of plausible values)
Psychometric analysis*

1) Psychometric quality of items was evaluated
2) Comparable scales across countries and assessments were established through IRT scaling
   constructs were similar enough to establish a concordance
3) Plausible value methodology was used to produce mathematics and literacy scores separately for each assessment
4) Concordance tables for mathematics and reading were established based on the mathematics and literacy scores of the source (ERCE/PASEC) and target assessments (TIMSS/PIRLS)

*Details of the analyses performed by the TIMSS and PIRLS ISC can be found here: https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html
Scale concordance

• Scale concordance establishes a relationship between scores on different assessments or tests that measure similar (but not identical) constructs

• Concordance tables allow inferences about likely score ranges on PIRLS or TIMSS scales – given an ERCE or PASEC score – to be drawn

• Score ranges account for a statistical measure of uncertainty in the estimation, due to factors like:
  – Differences in school curricula
  – Test language vs. language spoken at home
  – Cultural differences
  – Socioeconomic or sociodemographic differences
### Reading* concordance tables – extract

<table>
<thead>
<tr>
<th>ERCE Reading Score</th>
<th>Projected Score on PIRLS Scale</th>
<th>PASEC Reading Score</th>
<th>Projected Score on PIRLS Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD**</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>680</td>
<td>467</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>700</td>
<td>482</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>720</td>
<td>488</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>740</td>
<td>501</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

*Similar table exists for the TIMSS Mathematics Concordance

**SD=standard deviation
ERCE/PASEC link students reaching PIRLS benchmarks serving as MPLs*  

<table>
<thead>
<tr>
<th>ERCE country</th>
<th>PIRLS Low Benchmark (400)</th>
<th>Rosetta Test % (SE)</th>
<th>Concordance Table % (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colombia</td>
<td></td>
<td>85.9 (1.2)</td>
<td>83.6 (1.3)</td>
</tr>
<tr>
<td>Guatemala</td>
<td></td>
<td>63.2 (1.9)</td>
<td>67.3 (1.4)</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>74.5 (1.1)</td>
<td>75.5 (1.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PASEC country</th>
<th>PIRLS Low Benchmark (400)</th>
<th>Rosetta Test % (SE)</th>
<th>Concordance Table % (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burundi</td>
<td></td>
<td>4.1 (0.7)</td>
<td>10.5 (1-0)</td>
</tr>
<tr>
<td>Guinea</td>
<td></td>
<td>19.7 (2.1)</td>
<td>15.6 (2.2)</td>
</tr>
<tr>
<td>Senegal</td>
<td></td>
<td>41.0 (3.9)</td>
<td>36.3 (3.4)</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>21.6 (1.5)</td>
<td>20.8 (1.4)</td>
</tr>
</tbody>
</table>

*MPL = Minimum Proficiency Level

PIRLS Low benchmark -> serving as MPL for reading

Burundi: Differences between Test and Concordance likely due to late transition to French

Variability in single countries, but on average projections match very well
Proportion of students achieving MPLs

- Concordance tables allow for the generation of PVs on the PIRLS and TIMSS scales for students participating in regional assessments; this can be used to estimate the percentage of students that reach a given benchmark.
- Allows for the comparison of countries participating in regional and global assessments on the same scale.
Rosetta: global coverage for reading
Reading: proportion of students reaching MPLs
Rosetta: global coverage for mathematics
Mathematics: proportion of students reaching MPLs

Proportion of ERCE students achieving MPL for Mathematics

- **ERCE* G6 L3 (789 pts)**

Proportion of PASEC students achieving MPL for Mathematics

- **PASEC* G6 L3 (609 pts)**

Proportion of TIMSS students achieving MPL for Mathematics

- **TIMSS* G4 Interm. (475 pts)**
Note of caution when comparing students reaching MPLs across studies

• MPLs are set independently by each assessment, depending on context and focus

• Examples for benchmarks serving as MPLs for reading:
  – ERCE L3 (G6): “interpret expressions in figurative language based on clues that are implicit in the text”
  – PASEC L4 (G6): “combine their decoding skills and their mastery of the oral language to grasp the literal meaning of a short passage”
  – PIRLS low (G4): “When reading predominantly simpler Literary Texts, students can: locate and retrieve explicitly stated information, actions or ideas/Make straightforward inferences...”
• Rosetta is an important milestone in the development of global indicators that can be compared over countries and time

• Rosetta implements a resource-saving approach, allowing already-existing measurement tools to be statistically connected

• The project gave scientific evidence that the concordance approach allowed for the establishment of comparable scales for assessments that measure similar but somewhat different constructs
  – Rosetta is the only project that quantifies the associated uncertainties

• Thus, the concordance allows countries to compare their students’ achievement globally and provides data to measure global progress towards SDG 4.1.1b

• Information from Rosetta and from the regional assessments complement each other; together, they improve the quality and relevance of the information available
Outlook

- While scores vary more using the concordance tables when compared to a measurement with the same assessment, it is the statistically soundest procedure to allow for comparisons between different stand-alone assessments.
- Larger national sample sizes and adding more countries to the project could stabilize concordance further.
- To get higher global coverage, a comparable concordance should be extended to further regional assessments.
- Likewise, the approach may be extended to countries’ national assessments.
Thank you for your attention!

Oliver Neuschmidt (IEA)
Oliver.Neuschmidt@iea-hamburg.de

For more information see Rosetta Stone policy brief and analysis reports:
• https://tcg.uis.unesco.org/rosetta-stone/
• https://timssandpirls.bc.edu/Rosetta-Stone-Reports/