DEVELOPMENT AND VALIDATION OF THE RUSSIAN VERSION OF THE IPIPS FIRST GRADERS STUDY

Alina Ivanova
National Research University
Higher School of Economics
Moscow, Russia
Russian education: primary level

- **45,107 preschool institutions across Russia**
- **Not compulsory**
- **The majority of children do attend**

- **1,500 000 first graders in Russia**
- **1st of September – day X**
- **Children start school at the age of 7**

The primary education in Russia is regulated by Federal state educational standard. The Standard includes requirements regarding:

- learning outcomes
- core curriculum
- conditions for implementing the core curriculum
Background

No standardized LSA at the primary school level in Russia

Author tests and diagnostics

Assessment practices throughout the world

EDI (Early Development Instrument), Canada

Monitoring and Evaluation system for Primary Schools, The Netherlands

PIPS (Performance Indicators in Primary Schools), UK
The Russian iPIPS version

- **Writing**
- **Vocabulary**
- **Phonological awareness**
  - Repeating words
  - Rhyming words
- **Reading**
  - Ideas about Reading
  - Letters
  - Words
  - Reading short story
  - Reading passages
- **Mathematics**
  - Counting
  - Simple sums
  - Digits
  - Number manipulations
  - Math problems

**Behavior survey**
**PSED survey**
Test adaptation includes many activities from deciding whether or not a test could measure the same construct in a different language and culture, to checking equivalence of the initial and adapted test versions (Hambleton, 2005).
Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests . . . In the current validity conception, different forms of evidence on the validity of tests should not be considered to represent distinct types of validity, but validity should be considered a “unitary concept” (American Educational Research Association et al., 1999, p. 9-11).

The Dutch rating system was chosen as a basis for conducting the validity study (Evers, A., 2010).
## Validation study: Math scale

### Description: Russian math scale

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sums</td>
<td>Simple sums presented informally using pictures</td>
</tr>
<tr>
<td>Number</td>
<td>Number manipulation – the child is asked how many more or less a number is than a target</td>
</tr>
<tr>
<td>Sums A</td>
<td>Sums – addition and subtraction problems presented without symbols</td>
</tr>
<tr>
<td>Sums B</td>
<td>More difficult mathematics problems including sums presented with formal notation</td>
</tr>
</tbody>
</table>

### Sample

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>% of female</td>
<td>49%</td>
<td>50.4%</td>
</tr>
<tr>
<td>% of comprehensive schools</td>
<td>62%</td>
<td>56%</td>
</tr>
<tr>
<td>% of urban schools</td>
<td>72%</td>
<td>100%</td>
</tr>
<tr>
<td>in total</td>
<td>310</td>
<td>2741</td>
</tr>
</tbody>
</table>
Dimensionality study

**Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)**

<table>
<thead>
<tr>
<th></th>
<th>Empirical</th>
<th>Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total raw variance in observations</td>
<td>67.1 100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Raw variance explained by measures</td>
<td>24.1 42.8%</td>
<td>41.3%</td>
</tr>
<tr>
<td>Raw variance explained by persons</td>
<td>10.1 17.7%</td>
<td>17.3%</td>
</tr>
<tr>
<td>Raw variance explained by items</td>
<td>14.0 24.6%</td>
<td>24.0%</td>
</tr>
<tr>
<td>Raw unexplained variance (total)</td>
<td>33.0 57.7% 100.0% 56.7%</td>
<td></td>
</tr>
<tr>
<td>Unexplained variance in 1st contrast</td>
<td>1.7 2.8%  5.1%</td>
<td></td>
</tr>
<tr>
<td>Unexplained variance in 2nd contrast</td>
<td>1.5 2.6%  4.5%</td>
<td></td>
</tr>
</tbody>
</table>

**Scree Plot**

**Evidence for fair test use: cross-regional comparison**

**Math scale**
The psychometric quality of the Math scale

- 33 items math scale
- Person reliability 0.87
- Person separation index 2.57
The predictive validity of Math scale

- 310 children of the Velikiy Novgorod sample:
  - Their results on the iPIPS math scale
  - Their school grades in the beginning of the 2nd and of the 3d grades

The correlation of children’s iPIPS math scores and their school math grades

<table>
<thead>
<tr>
<th>Spearman correlation</th>
<th>2nd grade of school, math</th>
<th>3d grade of school, math</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPIPS: math</td>
<td>.510**</td>
<td>.471**</td>
</tr>
</tbody>
</table>

* p <0.05, ** p <0.001

The content validity

10 independent experts (psychology, pedagogy, educational science, educational policy) confirmed the feasibility of the iPIPS instrument for assessment of children starting school and for measurement of their progress during the first year of schooling.
Validation study: Behavior survey

Behaviour rating scale

Based on the modified (for young children in classroom) DSM-IV diagnostic criteria for ADHD

21 items

Two subscales:
- Inattentiveness
- Impulsivity/Hyperactivity

<table>
<thead>
<tr>
<th>Sample</th>
<th>Krasnoyarsk, 2014-2015</th>
<th>Moscow, 2015-2016*</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of female</td>
<td>49%</td>
<td>48%</td>
</tr>
<tr>
<td>% of comprehensive schools</td>
<td>63%</td>
<td>100%</td>
</tr>
<tr>
<td>% of urban schools</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>1303</td>
<td>2625</td>
</tr>
</tbody>
</table>
Dimensionality study

Preliminary analysis indicated at least two dimensions:

- Related to inattentiveness
- Related to impulsivity/hyperactivity
Theoretical models

Behavior*
- Inattentiveness
- Impulsivity/Hyperactivity

Behavior**
- Inattentiveness
- Impulsivity
- Hyperactivity

Behavior***
- Inattentiveness
- Impulsivity/Hyperactivity
  - Impulsivity
  - Verbal hyperactivity
  - Physical hyperactivity

Behavior rating scale

3-dimensional model

<table>
<thead>
<tr>
<th></th>
<th>Reliability</th>
<th>Impulsivity</th>
<th>Hyperactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulsivity</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>0.89</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>0.95</td>
<td>0.79</td>
<td>0.65</td>
</tr>
</tbody>
</table>

2-dimensional model

<table>
<thead>
<tr>
<th></th>
<th>Reliability</th>
<th>Impulsivity/ Hyperactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulsivity/Hyperactivity</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>0.94</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Convergent validity study

ADHD rating scale (Suhotina, Egorova, 2008)  
- 18 items, 3 sub-scales
  - Impulsivity
  - Hyperactivity
  - Inattentiveness

Pearson correlation

<table>
<thead>
<tr>
<th></th>
<th>Imp./Hyper. iPIPS</th>
<th>Inattentiveness iPIPS</th>
<th>Impulsivity</th>
<th>Hyperactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inattentiveness iPIPS</td>
<td>.798**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impulsivity</td>
<td>.876**</td>
<td>.690**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.884**</td>
<td>.667**</td>
<td>.987**</td>
<td></td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>.766**</td>
<td>.814**</td>
<td>.882**</td>
<td>.854**</td>
</tr>
</tbody>
</table>

**p<0.01

Sub-sample
Moscow 2015-2016
723, 1-graders
The predictive validity of the behavior rating scale

- 1303 children of the Krasnoyarsk sample:
  - Their results on the iPIPS math and reading scale at the end of the first school year
  - Their scores on the Behavior rating scale

The correlation of children’s iPIPS math and reading scores and their behavior rating scale scores

<table>
<thead>
<tr>
<th>Pearson correlation</th>
<th>Impulsivity/ Hyperactivity</th>
<th>Inattentiveness</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inattentiveness</td>
<td>0.790**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>-0.199**</td>
<td>-0.384**</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>-0.120**</td>
<td>-0.305**</td>
<td>0.548**</td>
</tr>
</tbody>
</table>

* p <0.05, ** p <0.001

The content validity

3 independent experts (clinical psychology) confirmed the feasibility of the iPIPS Behavior rating scale for usage by primary school teachers for children’s potential behavioral problems identification
Conclusions

☑ Cognitive scales: confirmed
  - Psychometric quality of the scales
  - Association with children’ school grades (end of the 2nd and 3d school year)
  - Expert evaluation of the instrument content

☑ Behavior rating scale: confirmed
  - Psychometric quality of the scale
  - Association with children’ math and reading results on iPIPS end of year assessment
  - Correlation with similar constructs
  - Expert evaluation of the scale
Future steps

- Investigate the association between the iPIPS scales’ scores and children results of the assessment monitoring at the end of their 4th school year
- International comparative research of children iPIPS results from different countries and cultures
- Validation study of the PSED survey
Thank you