Baseline Assessment in Schools: The iPIPS project

Providing high quality value added information on school and system effectiveness

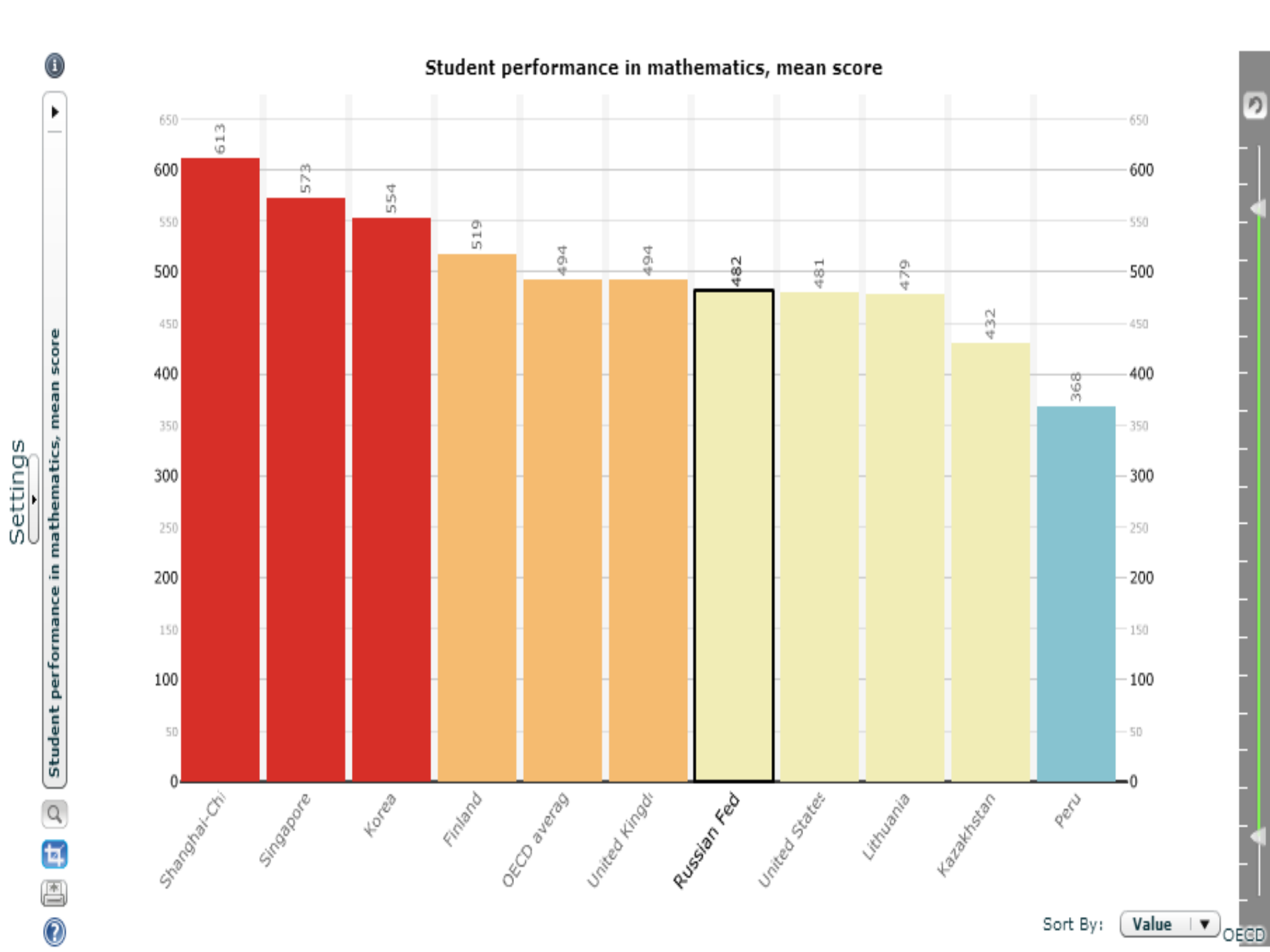
Professor David Hawker, Durham University
Elena Kardanova, Higher School of Economics

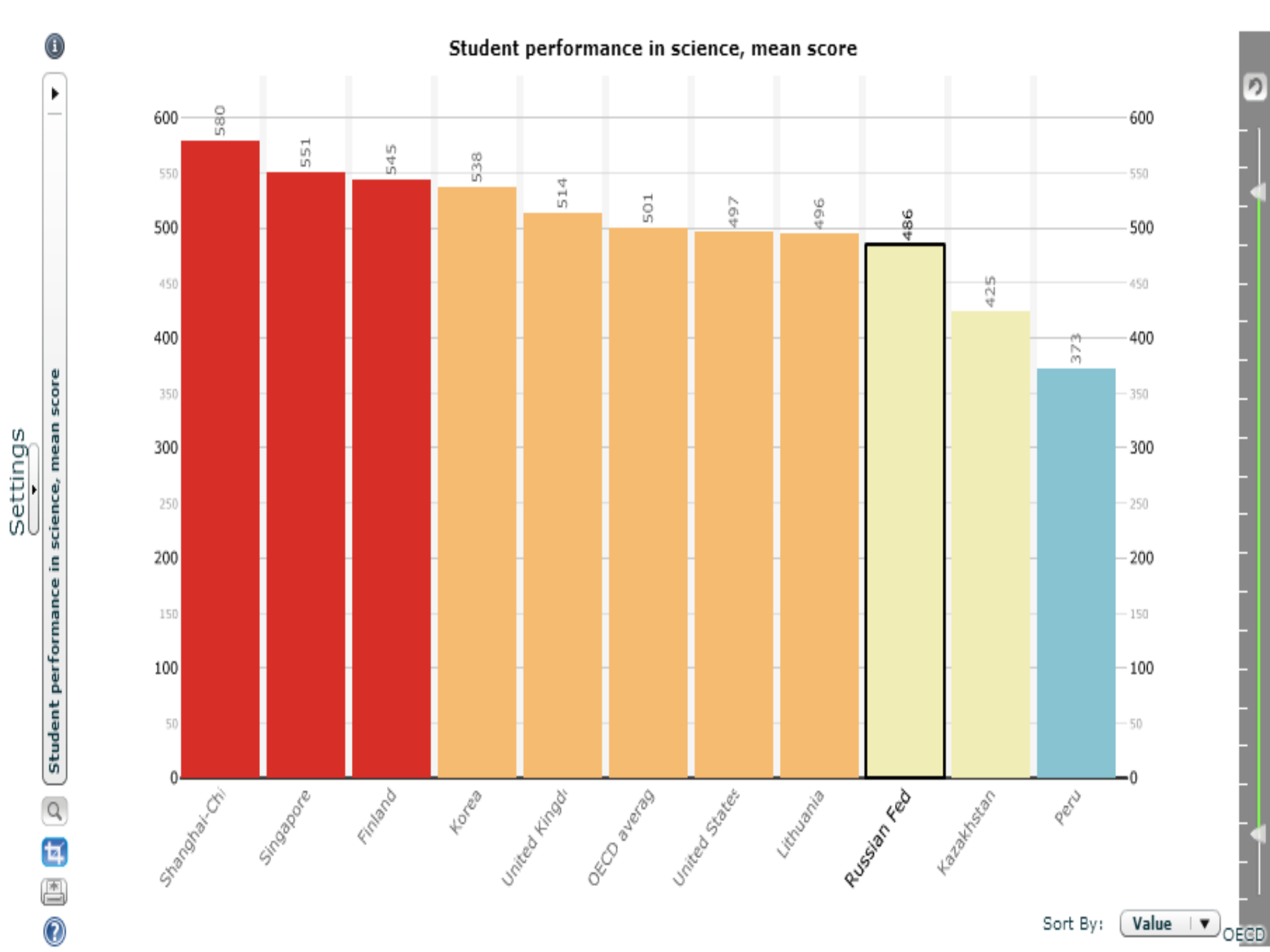
Outline of presentation

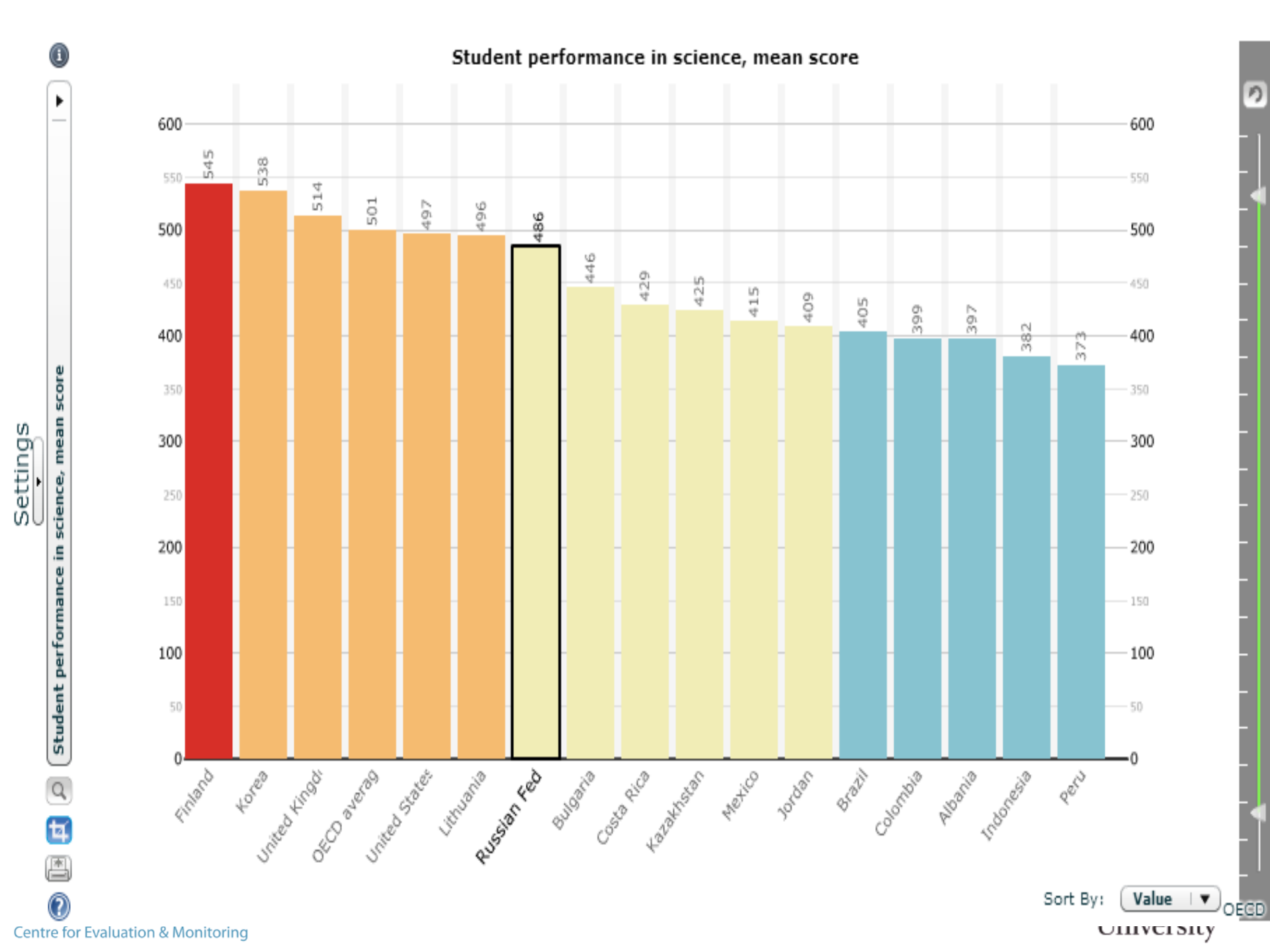
- Baseline assessment and value added in the international context: what questions need answering?
- Findings from research using PIPS
- Outline of the iPIPS project
- The Russian iPIPS trial
- Future plans

Settings

Student performance in reading, mean score


Student performance in mathematics, mean score
Settings
Student performance in mathematics, mean score
650
600
550
500
450
400
350
300
250
200
150
100
50
0
613
573
554
519
494
494
482
481
479
432
368
Shanghai-Chi
Singapore
Korea
Finland
OECD averag
United Kingdi
Russian Fed
United State
Lithuania
Kazakhstan
Peru
Sort By:
Value
OECD


Student performance in science, mean score
Settings
Student performance in science, mean score
600
550
500
450
400
350
300
250
200
150
100
50
0
580
551
545
538
514
501
497
496
486
425
373
Shanghai-Chi
Singapore
Finland
Korea
United Kingdi
OECD averag
United State
Lithuania
Russian Fed
Kazakhstan
Peru
Sort By:
Value
OECD


Student performance in science, mean score
Settings
Student performance in science, mean score
600
550
500
450
400
350
300
250
200
150
100
50
0
545
538
514
501
497
496
486
446
429
425
415
409
405
399
397
382
373
Finland
Korea
United Kingdi
OECD averag
United State
Lithuania
Russian Fed
Bulgaria
Costa Rica
Kazakhstan
Mexico
Jordan
Brazil
Colombia
Albania
Indonesia
Peru
Sort By:
Value
OECD
Centre for Evaluation & Monitoring

Which areas of PISA policy analysis have been influential in national policy-making processes?

a. Assessment and accountability	29
b. Learning environment	13
c. Early childhood education	13
d. Resource investment and allocation	12
e. Student selection and tracking	11
f. Governance (e.g. autonomy, choice, private/public).	11

OECD Working Paper 71 (2012)

When do the differences start?

- At home, before children start school?
- In pre-school?
- At school?

Nobody knows!

Five key questions

1. What is the value for money and relative effectiveness of different early years programs?
2. How much do children learn in their first year at school and how effective is the teaching?
3. How do different factors influence children's learning?
4. How can teachers and schools improve?
5. What are the best policies for long term effectiveness in children's learning?

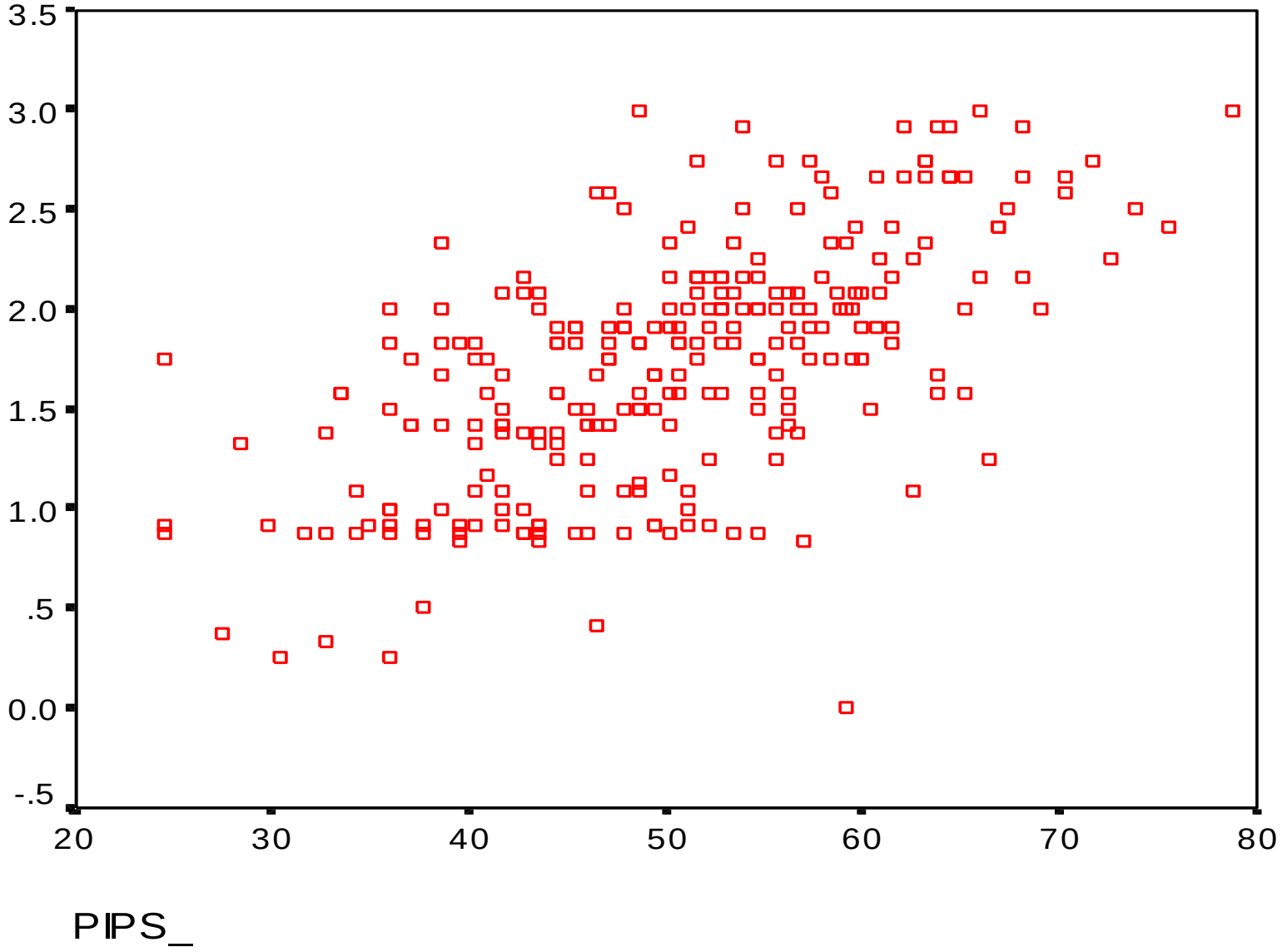
How will baseline assessment help?

1. ECEC Value for money – it will highlight differences in children's starting points, and relate these to their pre-school experience
2. Learning in the first year of school – it will measure children's progress by running the same assessment at the end of the year
3. Influence of different factors – it will relate data on a range of external and internal factors to children's learning and progress
4. Pedagogical improvement – it will provide diagnostic information to schools, and comparative information to policy makers, highlighting differences in practice between successful systems and less successful ones
5. Long term effectiveness – it will provide a baseline for later assessments: the early starters don't necessarily win in the end.

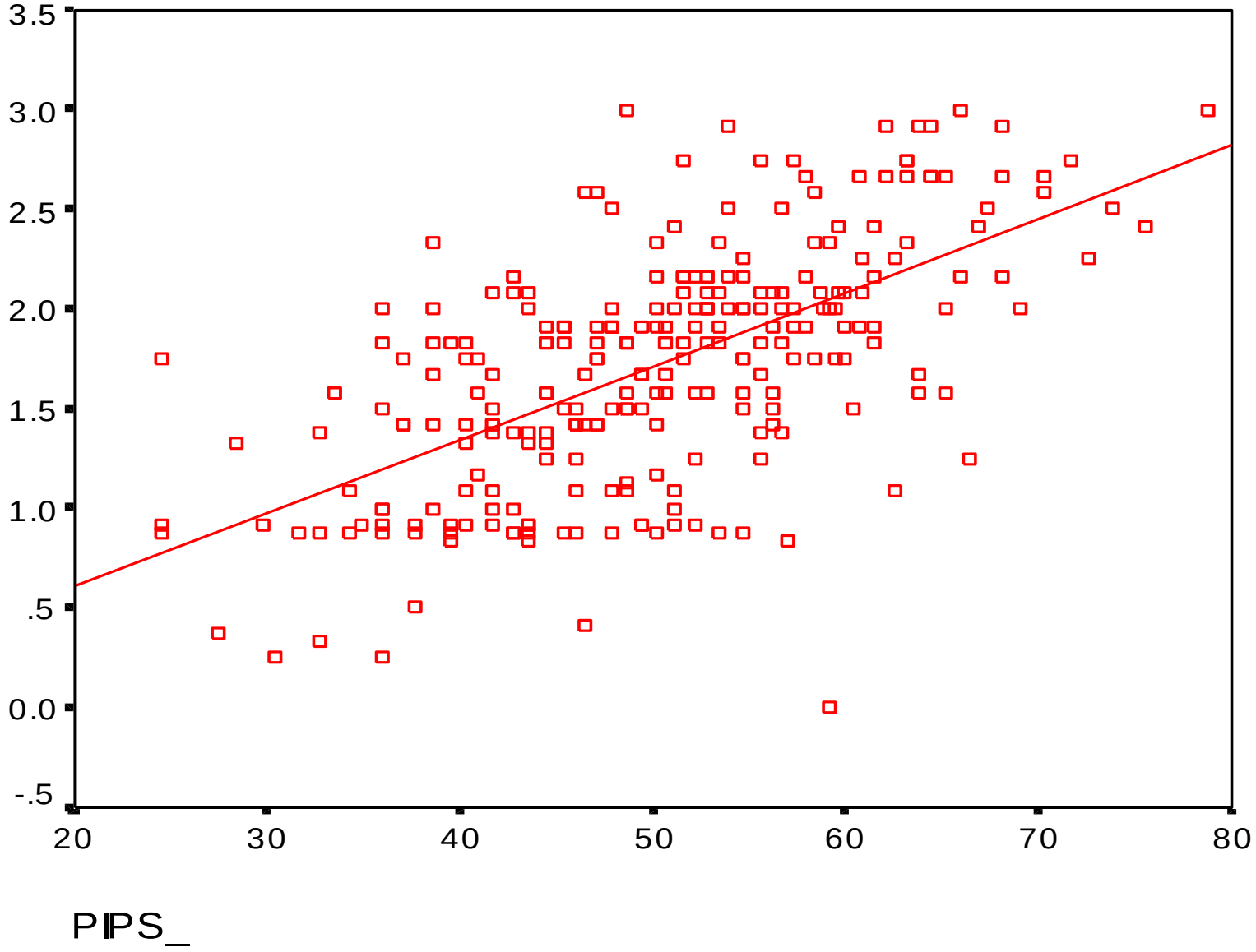
What is PIPS?

- A monitoring project started in the UK in 1994
- Baseline and follow up assessments
- Computer adaptive, child friendly
- High reliability (0.98 test-retest; 0.92 Cronbach's alpha)
- Good predictive validity (0.6-0.7 correlations to performance at age 7 and 11)
- Provides diagnostic feedback to schools, as well as data for policy makers
- Used in school effectiveness research in UK and elsewhere

Value added based on relative progress



Value added based on relative progress



Some research findings using PIPS: (1) The first year in school

- The two most powerful factors in children's achievement at the end of their first year in school are their prior achievement (effect size 2.5) and the school they attend (effect size 1.7).
- Attendance at pre-school has an effect size of only 0.3.
- So effective teaching in the first year is crucial to children's success.

Tymms, Merrell, & Henderson (1997). The First Year at School: A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation,* 3(2), 101-118

Some research findings using PIPS: (2) The first three years in school

- The effect of having good teachers for the first three years of school is 0.82 SD (ie large!)
- Younger children in the year group start at a disadvantage but catch up in the first 3 years
- Children who attended pre-school start with an advantage, but this does not increase as they get older

Tymms, Merrell and Henderson, 2000. Baseline Assessment and Progress during the First Three Years at School. *Educational Research and Evaluation* 6(2) p105 – 109.

Some research findings using PIPS: (3) The first seven years at school

- Over the first 7 years at school, most of the variance in children's achievement is explained by their prior achievement year on year
- Effective teaching in the early years has the most impact
- Effectiveness is almost entirely due to the teacher in the classroom – there is very little additional school effect.
- The effect size of having a poor teacher three years in a row is half a Standard Deviation – so the average child would slip to the bottom 16% in that time.

Tymms, Jones, Alborne and Henderson, 2009. The first seven years at school. *Educational Assessment and Evaluation Accountability*, 21, 67-80

Some research findings using PIPS: (4) Monitoring national performance over time

Between 2001 and 2008 there was very little change in the average levels of academic development of children starting school in England, despite massive government investment in pre-school education

Merrell and Tymms, 2011. Changes in Children's Cognitive Development at the Start of School in England 2001-2008. *Oxford Review of Education*

CEM
Centre for Evaluation & Monitoring

Durham University

Some research findings using PIPS: (5) Evaluation of a curriculum reform

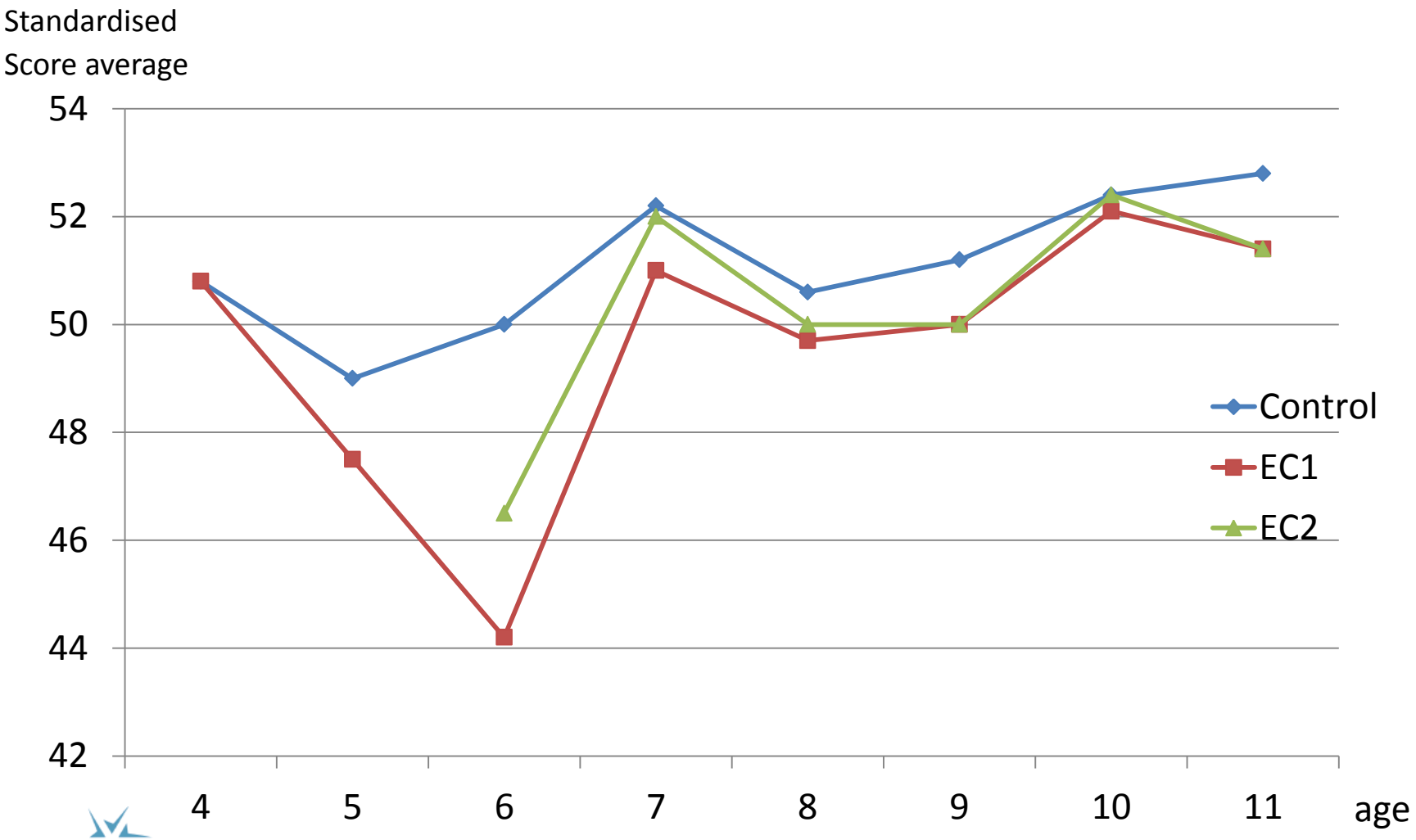
The introduction of a new play based curriculum for the first two years of school in Northern Ireland did not result in better progress in literacy and numeracy later on

McGuinness, Sproule, et al. (2013). "Impact of a play-based curriculum in the first two years of primary school: literacy and numeracy outcomes over seven years." *British Educational Research Journal*

CEM
Centre for Evaluation & Monitoring

Durham University

Impact of Enriched Curriculum in Northern Ireland on children's reading development



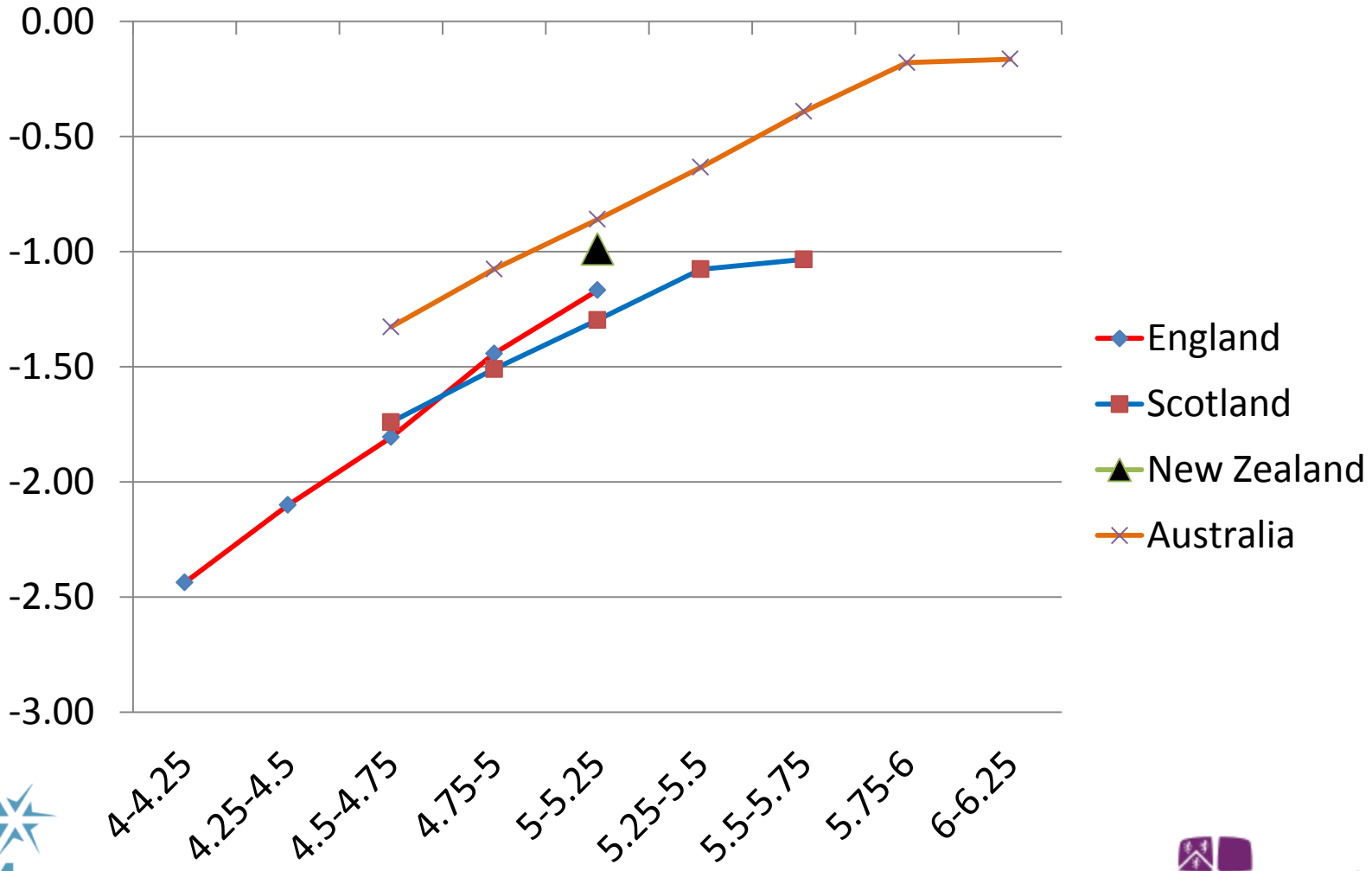
CEM
Centre for Evaluation & Monitoring

Durham University

Some research findings using PIPS: (6) Comparing children's performance across countries

Reading Development on entry

(Illustrative data– not fully representative)

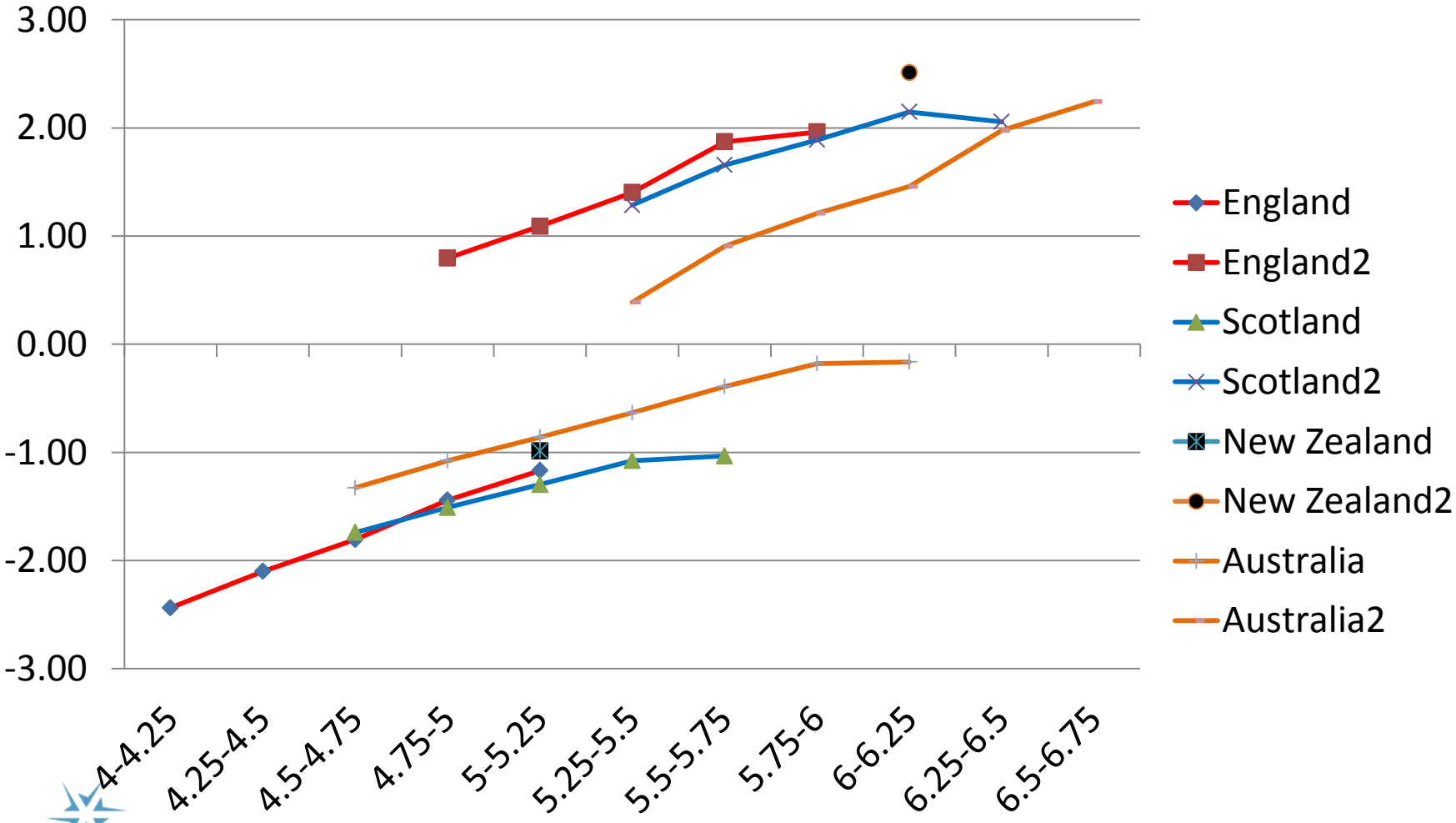


CEM
Centre for Evaluation & Monitoring

Durham University

Reading Development over the year

(Illustrative data– not fully representative)



CEM
Centre for Evaluation & Monitoring

Durham University

The iPIPS project

- An international monitoring survey of children starting school, using PIPS as the baseline and follow up assessment
- Designed to provide answers to the five key questions
- Partnership of research institutes, pooling expertise
- Links with OECD
- Information for policy makers, experts and schools

CEM
Centre for Evaluation & Monitoring

Durham University

iPIPS: What is Planned

- Adapt existing PIPS assessment specifically for international comparative use
- Sample based monitoring of 3000 children's developing abilities at start and end of first year in school per country/region
- International and country/regional analyses
- Data for schools to use diagnostically (not accountability or performance management)
- Pilots in 6-8 countries 2014-16
- To be offered more widely thereafter

CEM
Centre for Evaluation & Monitoring

Durham University

Russia: Why iPIPS?

- The object of the study: the range of children's skills and abilities, both cognitive and non-cognitive
- Individual assessment
- Standardized assessment with established psychometric properties and validity (on British sample), recognized in the world
- Special measurement technique lets evaluate an individual progress of a child over the first year
- Computer adaptive test
- Gentle and precise assessment of each child
- Unique for the Russian school system

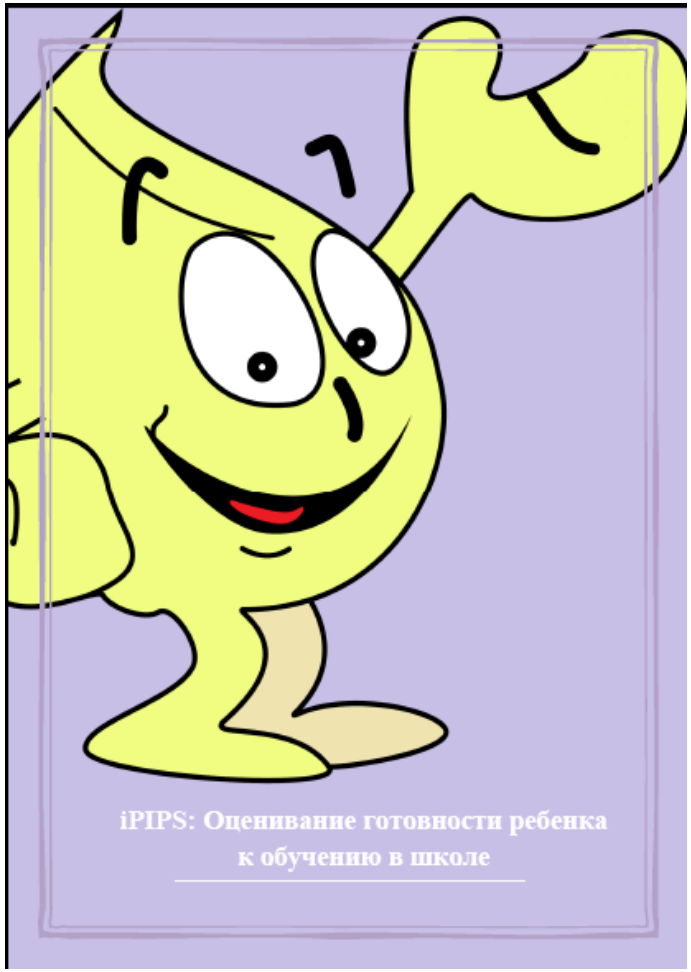
Adaptation of iPIPS in Russia

- Development of the Russian language test version of iPIPS
- Development of the procedures and carrying out two assessment cycles (testing children and collecting the context data), such as the baseline assessment and the first class end assessment
- The research on adaptation, including psychometric analysis of item properties comparing to the international database of iPIPS
- Validity study of the Russian version of iPIPS
- Development of the assessment method and a common scale construction for two cycles which is compatible with the international scale
- Primary data analysis

Center for Monitoring of the Quality of Education (Institute of Education, NRU-HSE) performs the work on adaptation and validization of iPIPS in the Russian Federation, according to the agreement with Durham University.

iPIPS structure

- A booklet with items for children (verbal part, math, attitudes)
- A questionnaire for parents
- A questionnaire for teachers:
  - Survey of social and emotional development of a child
  - Survey of teachers



Writing

27% of children can write down their full names correctly and with proper capitalization

У щенка красный мяч.

Щенку очень нравится играть с мячом на траве.

После игры ему хочется лечь и поспать.

Reading:
A short story

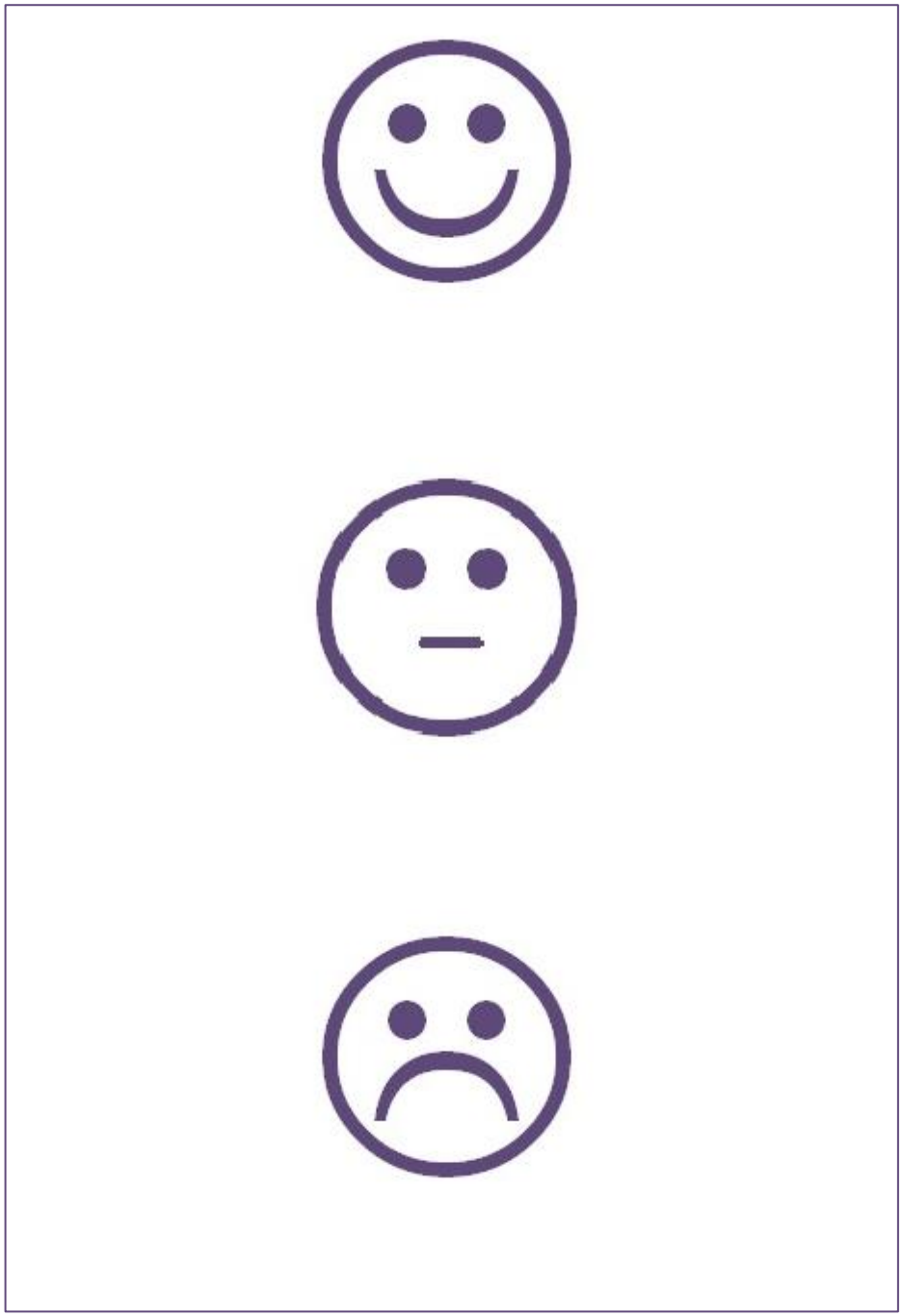
All word were read correctly: 31%
No words were read correctly: 17%

Examples of math tasks

1) Sasha wants to buy an orange which costs 12 rubles, which coins should he use?

2) What is 3 less than 7?

3) What is a half of 6?



Attitudes of a child

“Listening to stories"
“Looking at pictures in books".
“Drawing pictures"
“Building models out of Lego or any construction kit"
“Counting"
“Being at school"
“Playing out in the playground"
“Working on the computer"

Sample

Region: Velikiy Novgorod, Novgorod Region

Sample size: 311 first year students (5% of the population)

- Stratified random selection: a random selection of a settlement by given parameters, random selection of schools, classes and children in a class

Training the interviewers

- Selecting interviewers among students of the HSE master programme "Measurement in psychology and education"
- Training:
  - Getting familiar with iPIPS materials
  - Watching video of an example of assessment in the UK
  - Analysis of «difficult» situations during the assessment

Assessment procedure

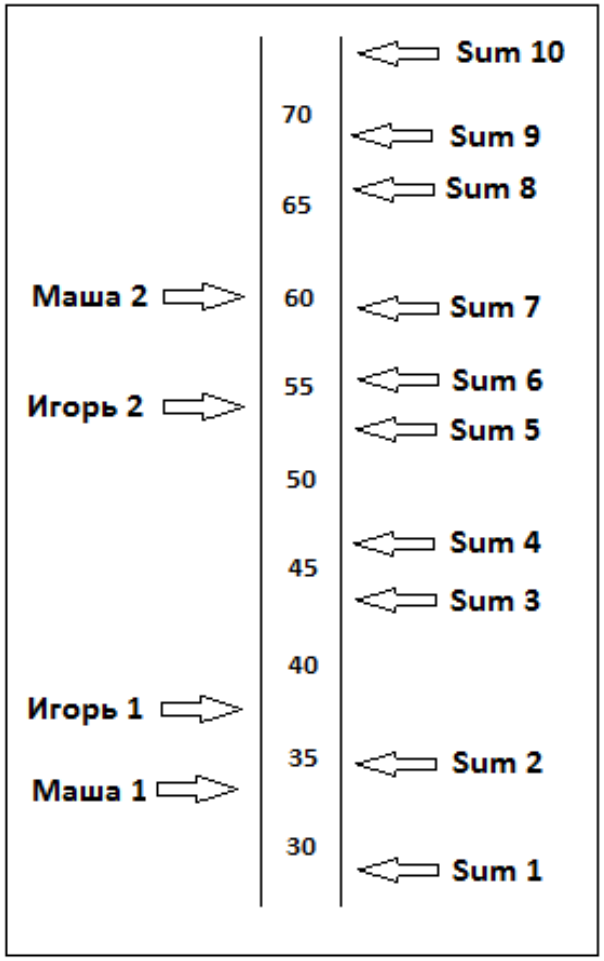
- Individual work with each child
- Duration of 20-30 min
- Play-based assessment
- A colourful booklet for a child
- A tablet computer with software for each interviewer

Diagnostic procedures

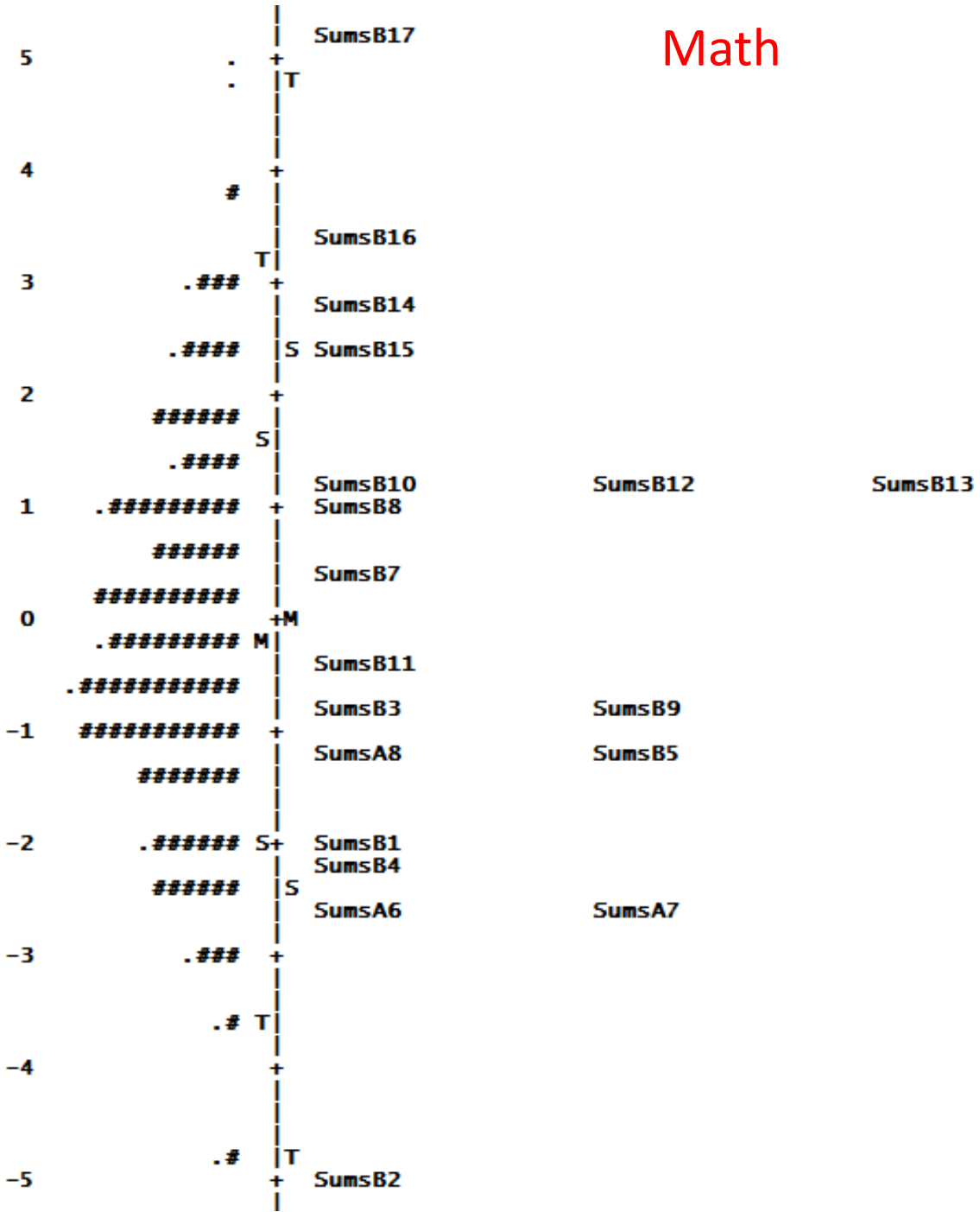
- Survey of parents (filling in a questionnaire)
- Survey of teachers
  - Survey of social and emotional development of a child
  - Teacher's survey

Data analysis: methodology

- Comprehensive approach based on modern principles of measurement in education.
- Item Response Theory (IRT) as a theoretic base.
- Three types of score interpretation: norm-referenced, criterion-referenced and self-referenced
- Family of Rash models was chosen for modeling
- The metric scale allows to compare results of different children and groups of pupils and to use various methods of math statistics to research the data and checking different hypotheses, equate the results of testing from different periods.

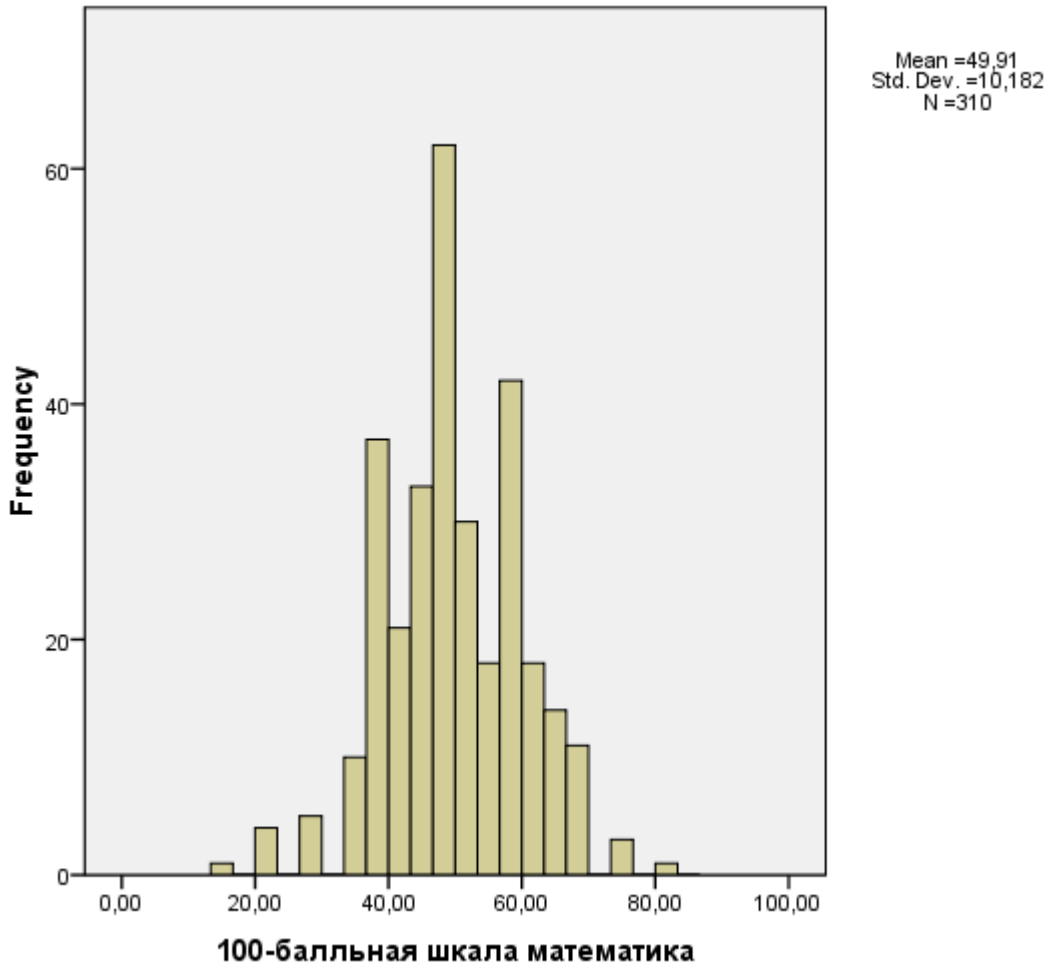


Math



sumsA6	Добавить к 2 щенкам еще два
sumsA7	Добавить к 3 каплям еще три
sumsA8	Добавить к 4 ракетам еще 4
sumsB1	Прикрыть половину велосипедов
sumsB2	Купить яблоко за 5 рублей
sumsB3	Число на один больше, чем 5?
sumsB4	Сколько будет 3 яблока и 2 яблока
sumsB5	7+3=
sumsB7	Число на 2 больше чем 6
sumsB8	Число на 3 больше чем 8
sumsB9	Купить апельсины за 12 рублей
sumsB10	Было 12 апельсинов, 4 отдал
sumsB11	14 людей в автобусе, на остановке зашли еще 3
sumsB12	Половина 6
sumsB13	Число на 3 меньше чем 7
sumsB14	Число на 6 меньше чем 15
sumsB15	17-15=
sumsB16	14+23=
sumsB17	Четверть 8

Students estimation

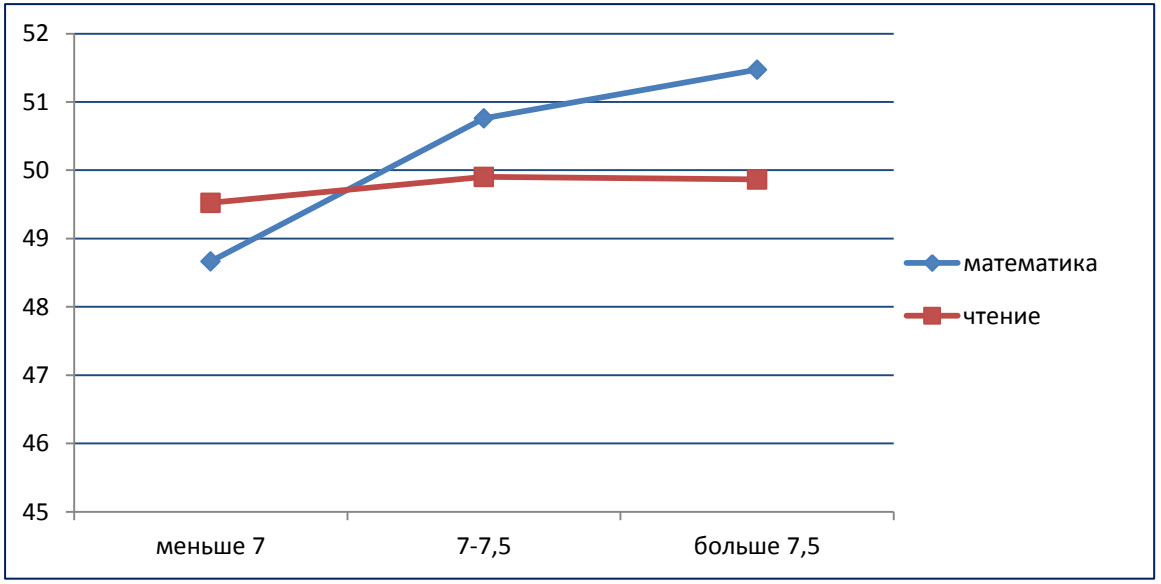


- Convenient 100 - point scale with mean=50 and SD=10

Primary analysis: results by age

Age	Math	Reading	Number of children
less than 7	48, 7	49,5	77
7-7,5	50,8	49,9	167
more than 7,5	51,1	49,7	43

No significant mean differences



Results of the primary analysis

- Child's gender:

  Math results of boys and girls do not differ, while in reading and phonetics girls' scores are slightly higher.

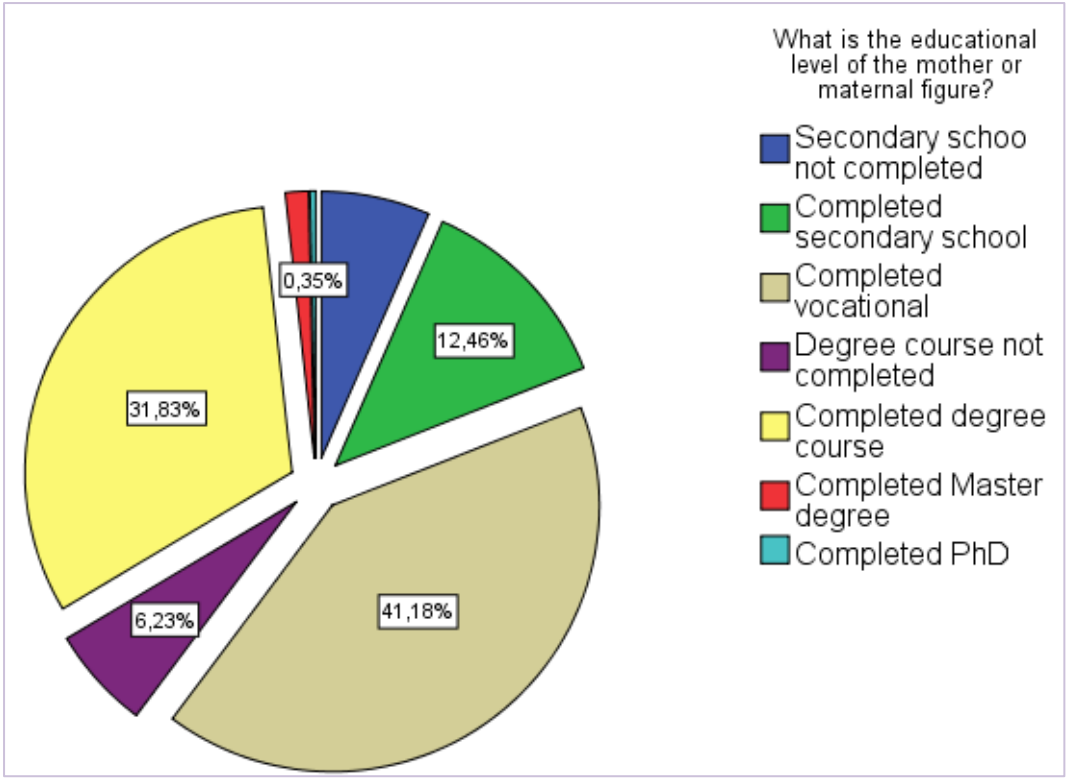
- Type of school:

  Students of schools with advanced study of some subjects get the highest scores in all tested areas, while the least scores were received by children from comprehensive schools. There are no differences between the results of children from gymnasiums and advanced study schools.

- Location:

  There are no differences in math scores between children from different settlements, but there are differences in reading scores.

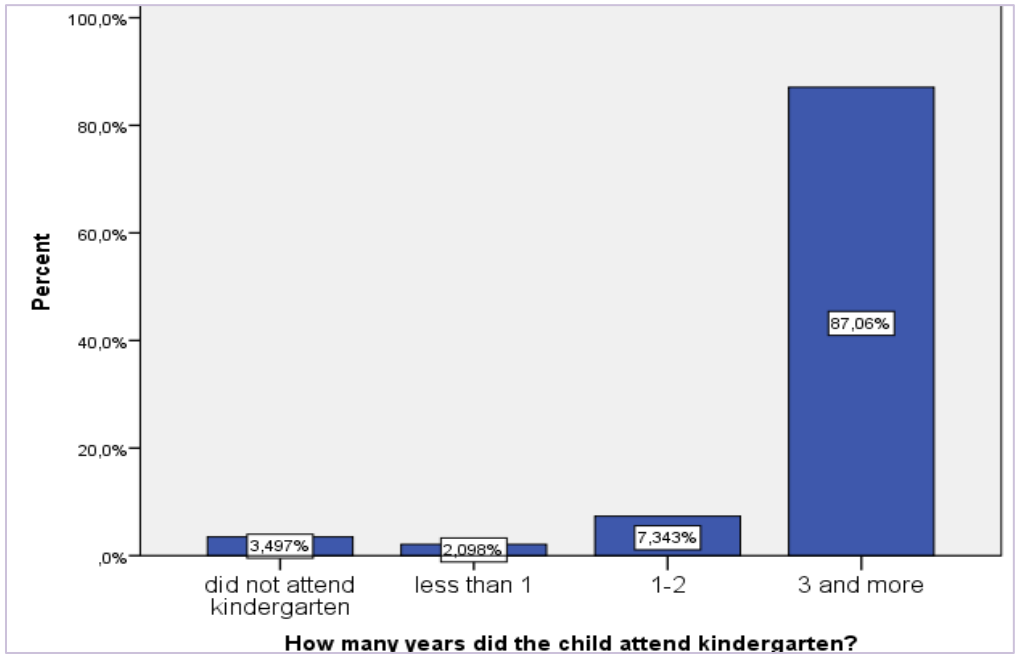
Educational level of parents



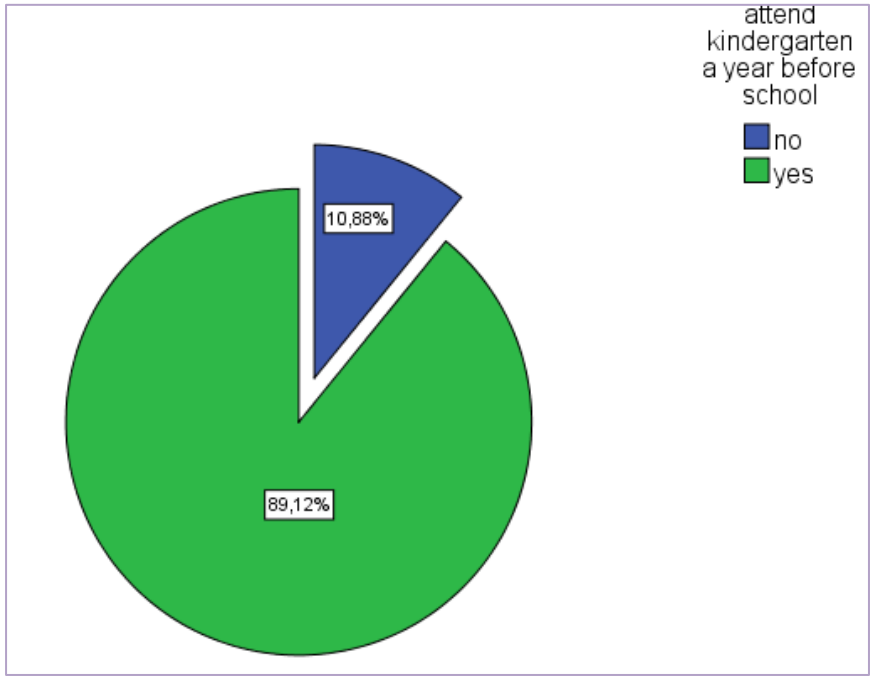
Children whose mothers **have higher education** read better and understand more in math than children whose mothers don't have higher education.

41% of mothers и 50% of fathers have vocational secondary education

t-test	Score difference
math	4.51*
reading	7.35*

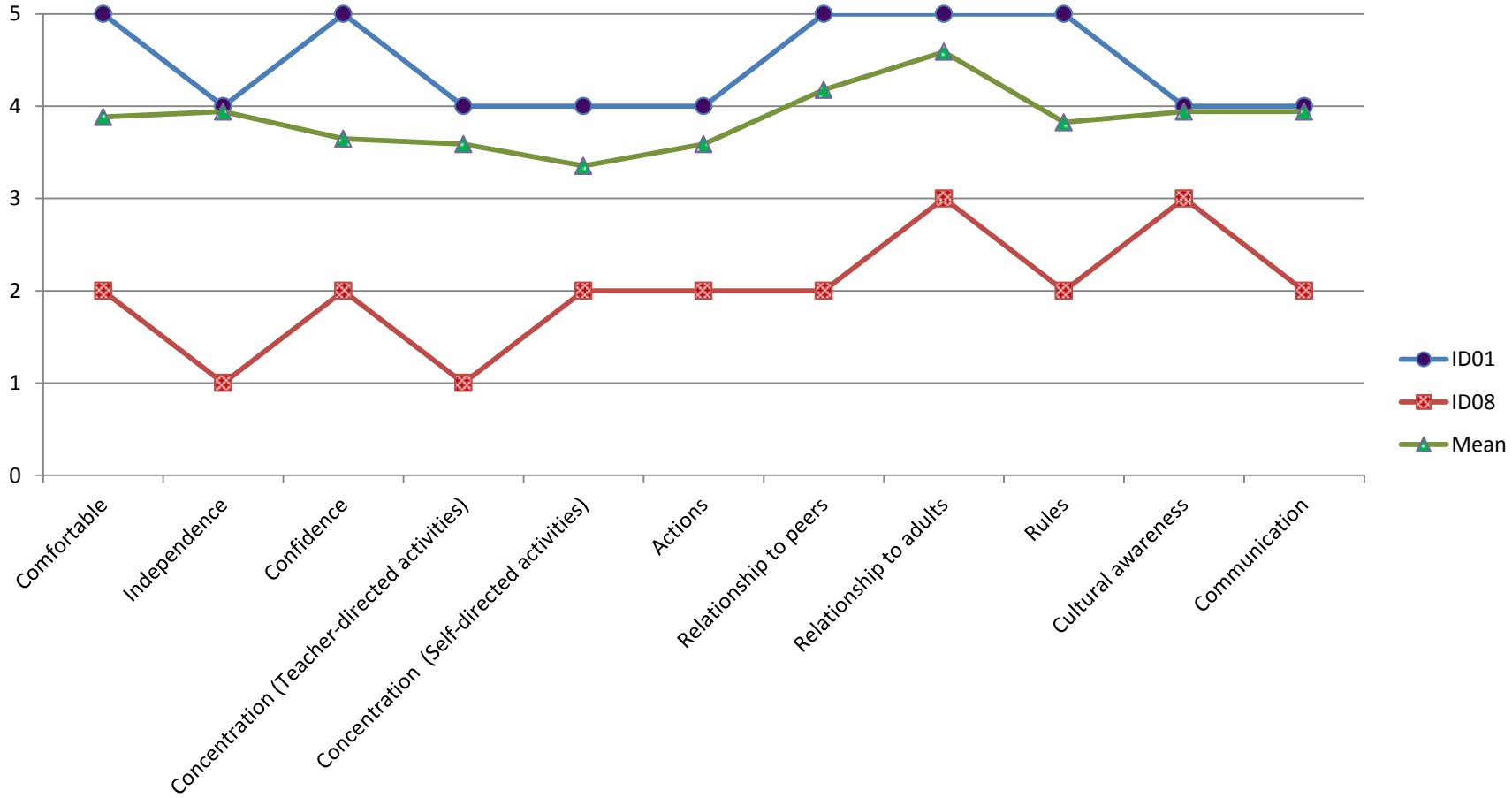


87% of children went to kindergarten for more than 3 years
89% went to kindergarten for one year just BEFORE school
63% went to some centers of development, sections and hobby-groups
55% went to special pre-school classes



28. Did a child go to kindergarten in the year just BEFORE SCHOOL?		N	Mean	Std. Dev.
Math	да	254	51	9.7
	нет	31	48	11.0
Reading	да	254	51	11.2
	нет	31	44	16.0

Profiles of social and emotional development of a child



Plans for 2014-2015 гг.

- Second stage of pilot testing (May 2014, the same sample)
  - Creating a common scale
  - Developing the methodology of individual progress assessment

- 1st research stage of main study in 2-3 Russian regions
  (September 2014, Sample of 3000 children)
  - Validization of the test
  - International comparative research of readiness to school

- 2nd research stage of main study
  (May 2015, the same sample of 3000 children)
  - Individual progress assessment
  - International comparative research of children's progress over the first school year

Report for a seminar «Assessment in social sciences»

David Hawker, Alina Ivanova

Investigating Validity of International and Cross-cultural assessment: iPIPS Study (first results)

Institute of Education ,

National Research University Higher School of Economics

February, 21, 2014, 16.00

(Milutinskiy., 13, room 406)

CEM
Centre for Evaluation & Monitoring

Durham University

Thank You

www.ipips.org

ЦМКО

CEM
Centre for Evaluation & Monitoring

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Durham University