

Measuring quality of engineering education in elite and non-elite universities in Russia and China

Elena Kardanova,

Ekaterina Enchikova,

NRU Higher School of Economics

Moscow, Russia

Prashant Loyalka

Stanford University , USA

- A major goal of university systems is to produce skilled graduates (Spellings, 2006; Alexander, 2000)
- Skilled graduates can contribute towards the productivity and innovation → higher economic growth (Goldin and Katz, 2008; Autor et al., 2003; Bresnahan et al., 2002; Bresnahan, 1999; Katz and Krueger, 1998)
- Failing to produce skilled graduates may hinder the capacity of countries to compete in the global knowledge economy → stifle growth (Hanushek and Woessman, 2012; Hanushek and Woessman, 2008)

Although there is high and increasing interest from researchers and policymakers, few studies have examined the quality of education at universities

- A couple of recent US studies show students are making modest gains in academic and higher order thinking skills (Pascarella et al., 2011; Arum and Roska, 2011)
- There are very few international comparison studies (AHELO)

ISHEL (International Study of Higher Education Learning) project

Two main goals:

1) Assess and compare university student skills (levels and gains) within and across countries

2) Examine which factors help students develop skills

- Focus on engineering students (Electrical Engineering and Computer Science) in Russia and China
- Assess skills over time
 - academic skills (math, physics based on our own tests)
 - major-specific skills (ETS computer science test, our own informatics test)
 - higher order skills (ETS new critical thinking & quantitative reasoning tests)

- A major barrier to assessing and comparing the skill gains of engineering students is the lack of valid, equitable, and cross-nationally comparable assessment instruments
- The purpose of this study is to prove reliability and cross-cultural comparability of test instruments for measuring the quality of engineering education in different types of institutions and countries.

Mathematics is taken for the presentation.

The requirements for test instruments

- Can be used to measure and compare
 - skill levels
 - skill gains
- Provide fair and reliable measurement of students' abilities
- Can be horizontally and vertically scaled to measure and compare absolute learning (gains) across and within countries

Steps for the development of valid, equitable, and cross-nationally comparable assessment instruments

1. Select comparable EE and CS majors across China and Russia
2. Select content and sub-content areas in math and physics (*with experts*)
3. Collect and verify items (*with experts*)
4. Conduct a pilot survey
5. Conduct a psychometric analysis

Step 1. Selecting comparable EE and CS majors across China and Russia

Since EE/CS are divided into many specialized majors

In China and Russia, we selected majors that have

- consistent coursework/curricula across universities within each country
- substantial overlap in courses and curricula across countries

Step 2. Selection of content areas for the tests

- content maps in math and physics that:
 - were based on the national curriculum standards for the sampled majors in Russia and China
 - covered areas that students should know at the start of their engineering studies (for the grade 1 test) and in the first two years of engineering studies (for the grade 3 test).

Test	Number of content areas	Number of sub-content areas
Mathematics, grade 1	8	36
Mathematics, grade 3	10	116
Physics, grade 1	10	49
Physics, grade 3	10	104

Experts surveying (stage 1)

- Interviewed 12 experts in each country (6 profs from elite institutions, 6 from non-elite institutions)
- Experts adjusted content maps to reflect what EE/CS students learn in math & physics (by grade 1 & by grade 3)

Test	Content areas selected based on expert evaluation
Mathematics, grade 1	<ul style="list-style-type: none"> ✓Functions and domains ✓Equations ✓Derivatives and their application ✓Mathematical reasoning and logic ✓Trigonometric functions and equations
Mathematics, grade 3	<ul style="list-style-type: none"> ✓Single variable differentiation ✓Single variable integration ✓Linear algebra ✓Multivariate differentiation ✓Series
Physics, grade 1	<ul style="list-style-type: none"> ✓Electromagnetic fields ✓Electromagnetic induction ✓Circuits ✓Optics ✓Oscillation and mechanical waves
Physics, grade 3	<ul style="list-style-type: none"> ✓Electricity and Electric Fields ✓Electromagnetic Induction ✓Magnetism and Magnetic Fields ✓Optics ✓Waves and Oscillation

Selection of content areas for the tests: results

Based on expert ratings of the importance of each content and sub-content area for the academic progress of students in EE and CS majors, we selected the content and sub-content areas with the highest average ratings in each subject

Step 3. Collection and verification of test items

- Collected tests items that fitted the content areas selected by experts for each subject and grade level
- The sources of test items: each country's past university entrance exams (China's Gaokao and Russia's Unified State Exam), other standardized exams and popular exercise books in both countries, ETS tests
- Requirements for item selection:
 - short and simple sentence structures and grammatical forms;
 - multiple-choice (MC) format (as it is the most familiar format for our target populations).

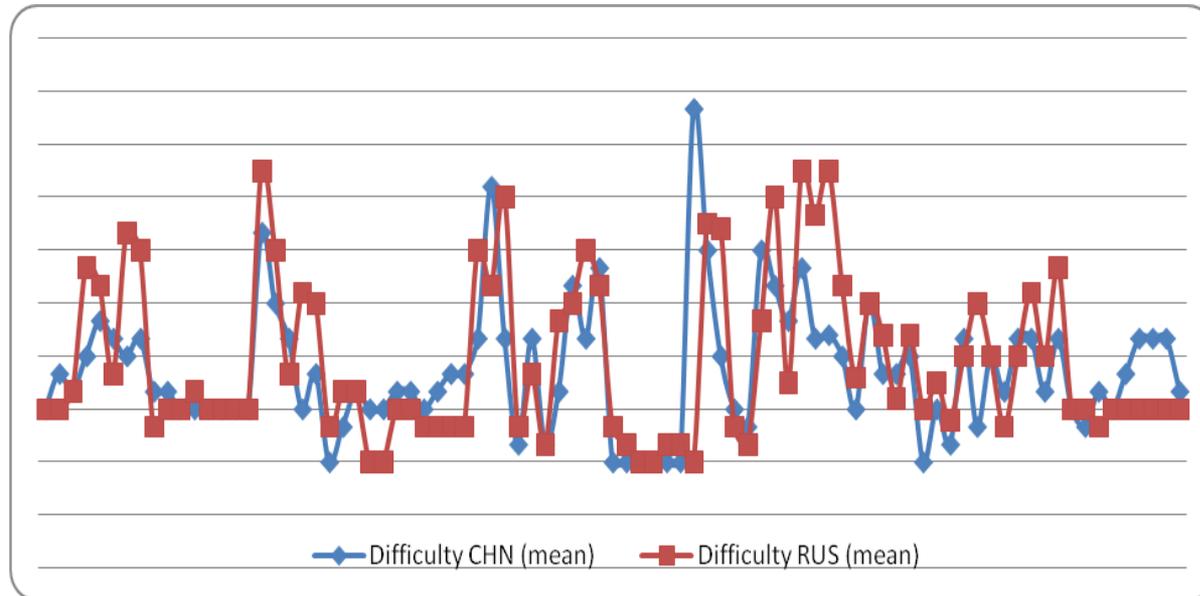
Experts surveying (stage 2)

We asked the same 12 experts from each country to rate the items according to four criteria:

- 1) comprehensibility of wording;
- 2) appropriateness in measuring the content area of interest;
- 3) expected difficulty;
- 4) expected time required to answer.

Results.

Item selection: consistence between the experts



- The Cronbach's Alpha coefficient was over 0.8 and correlations between the ratings were over 0.6 and significant at the .01 level
- There were statistically significant differences in the level of "severity" of the experts' evaluations. The "severity" of an expert was not related to nationality.
- Despite this difference in "severity," no experts demonstrated effects that might have affected the final ratings and brought bias into the evaluation procedure.

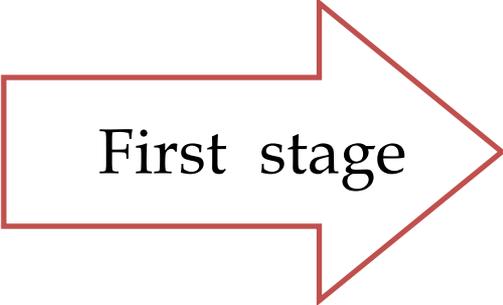
Step 4. Large scale piloting

- October 2014;
- 11 universities in China and 10 universities in Russia, both elite and non-elite universities, located both large and small cities across the country ;
- 1,797 students in China and 1,802 students in Russia;
- Paper and pencil format;
- Two 55-minute sessions (one for math and one for physics, random order)
- Students were also asked to fill out a short 10 minute questionnaire about their background and schooling experience after testing was completed.

Step 5. Ensuring psychometric quality and a common scale between grades and countries

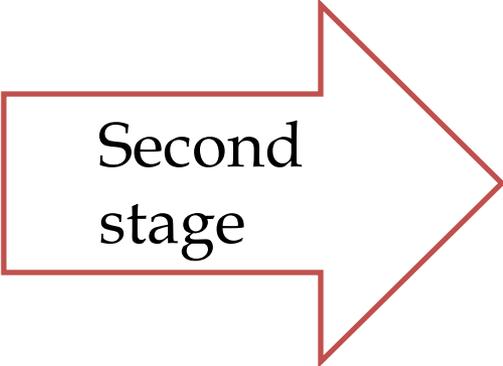
- Item Response Theory (IRT) modelling to conduct item analyses as well as tests of dimensionality and reliability.
- The one-parameter dichotomous Rasch model was used for the IRT analysis
- We also paid particular attention to differential item functioning (DIF) to provide evidence concerning the cross-national comparability of the test results and to ascertain the possibility of creating a common scale between the two grades and across the two countries.

Two stages of data analysis



First stage

Each grade analyzed separately to discover whether it would be possible to construct a common scale across countries within each grade.



Second stage

Two grades analyzed simultaneously, using common items included in both grades as a link. to determine whether it is possible to place all parameters onto a common scale between the two grades and across the two countries.

Results.

Psychometric analysis

Objective	Methodology
1. Choose “good” items	1. Classical Test Theory + IRT
2. Make sure we measure “ability” as a unidimensional construct	2. Test of dimensionality
3. Make sure items are fair across countries	3. DIF analysis
4. Look at the test as a whole: understand if the test is an appropriate measurement tool for the sample	4. Reliability study, variable map

Psychometric analysis

Objective	Methodology
1. Choose “good” items 2. Make sure we measure “ability” as a unidimensional construct	1. Classical Test Theory + IRT 2. Test of dimensionality
3. Make sure items are fair across countries	3. DIF analysis
4. Look at the test as a whole: understand if the test is an appropriate measurement tool for the sample	4. Reliability study, variable map

Goal: Make sure items are fair across countries
Approach: DIF analysis

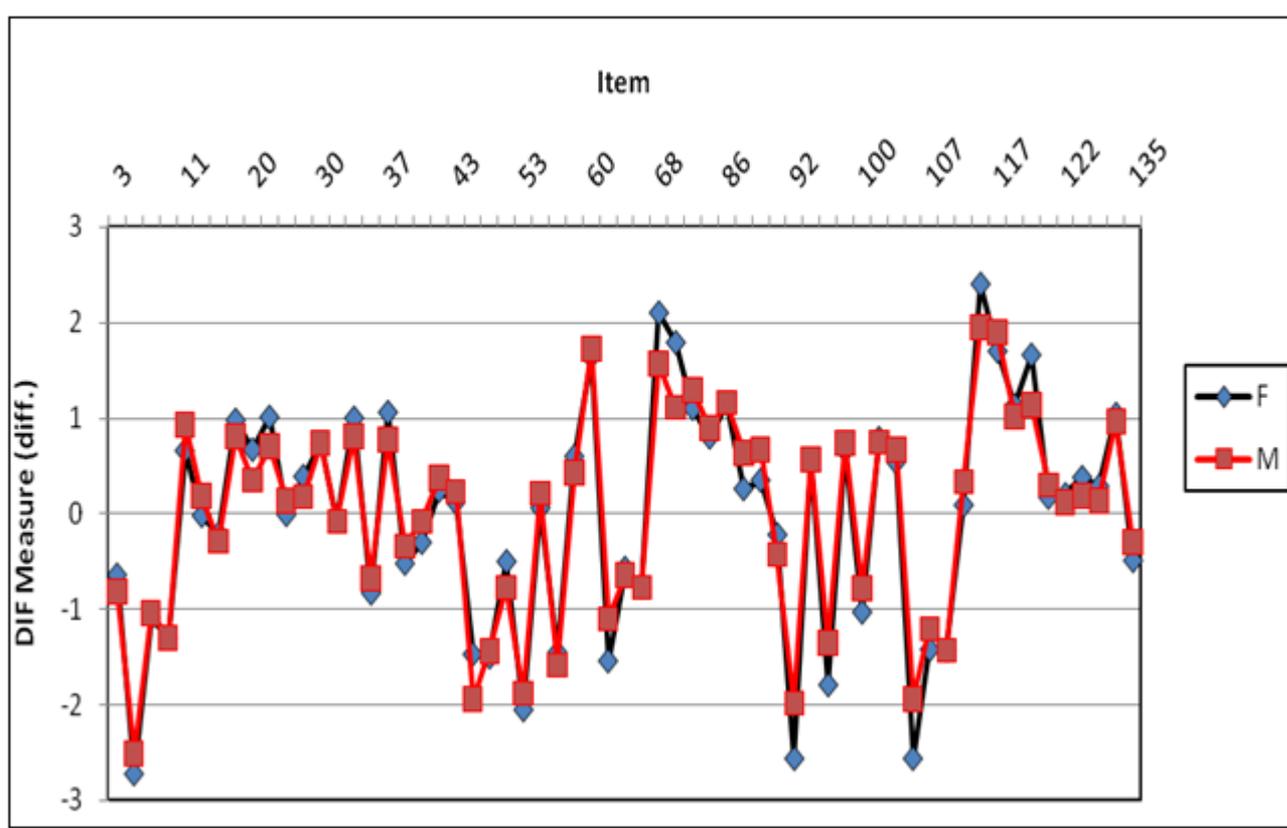
DIF (Differential Item Functioning):

when test participants with the same ability level who belong to different groups (e.g. gender, country) have varying chances to complete the item correctly.

So check at each item for any difference in performance by group, while controlling for ability.

- Mantel-Haenszel method
- ETS approach for DIF classification

Math test, grade 3, China + Russia, DIF across gender



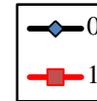
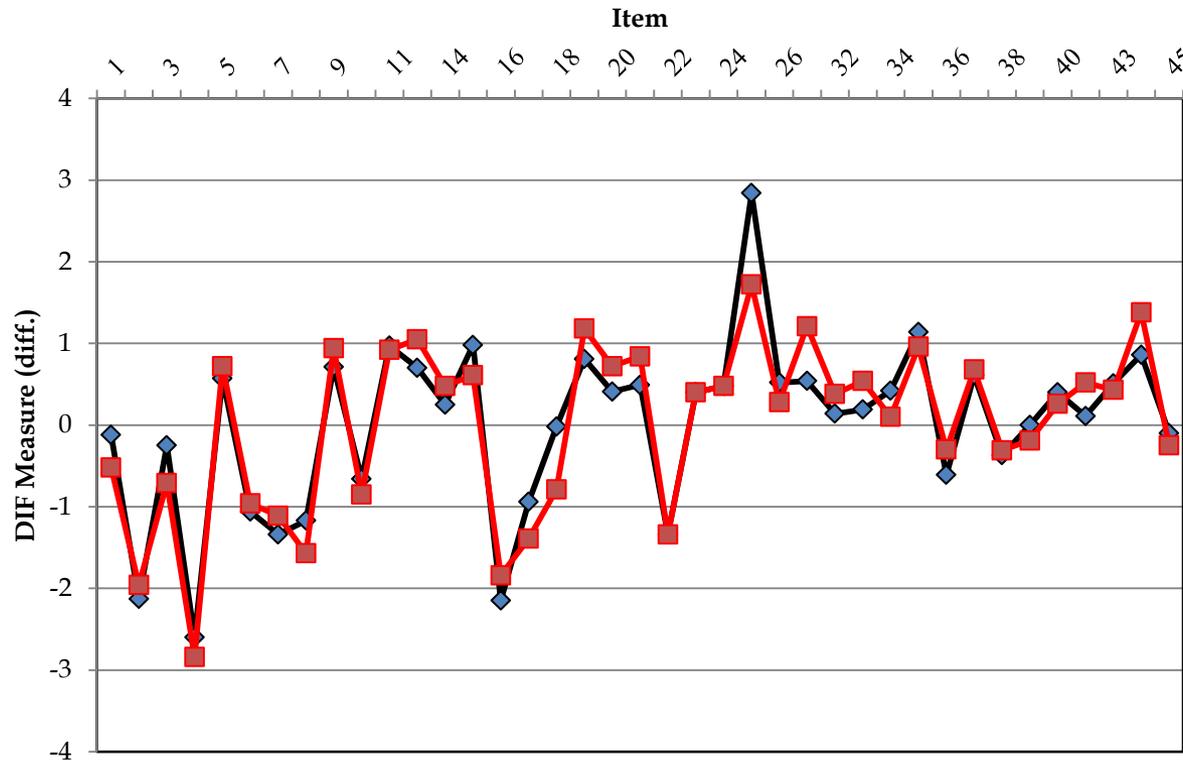
Females in blue, males in red.

There is no difference in item difficulty for boys and girls

Therefore,
NO DIF

No gender bias

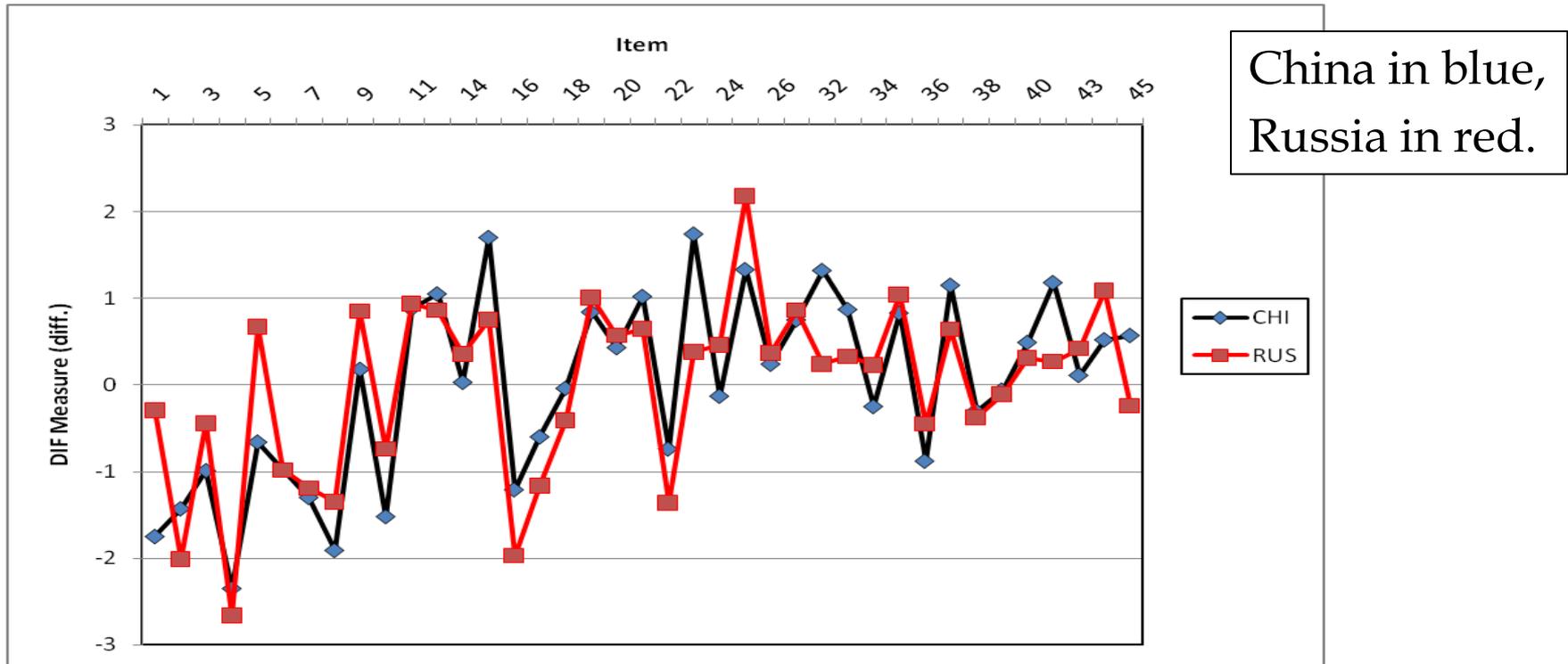
Math test, grade 3, Russia, DIF across elite and non-elite universities



Non-elite is blue,
Elite in red.

There is no significant
difference in item
difficulty

Math test, grade 3, Russia+China DIF analysis across countries



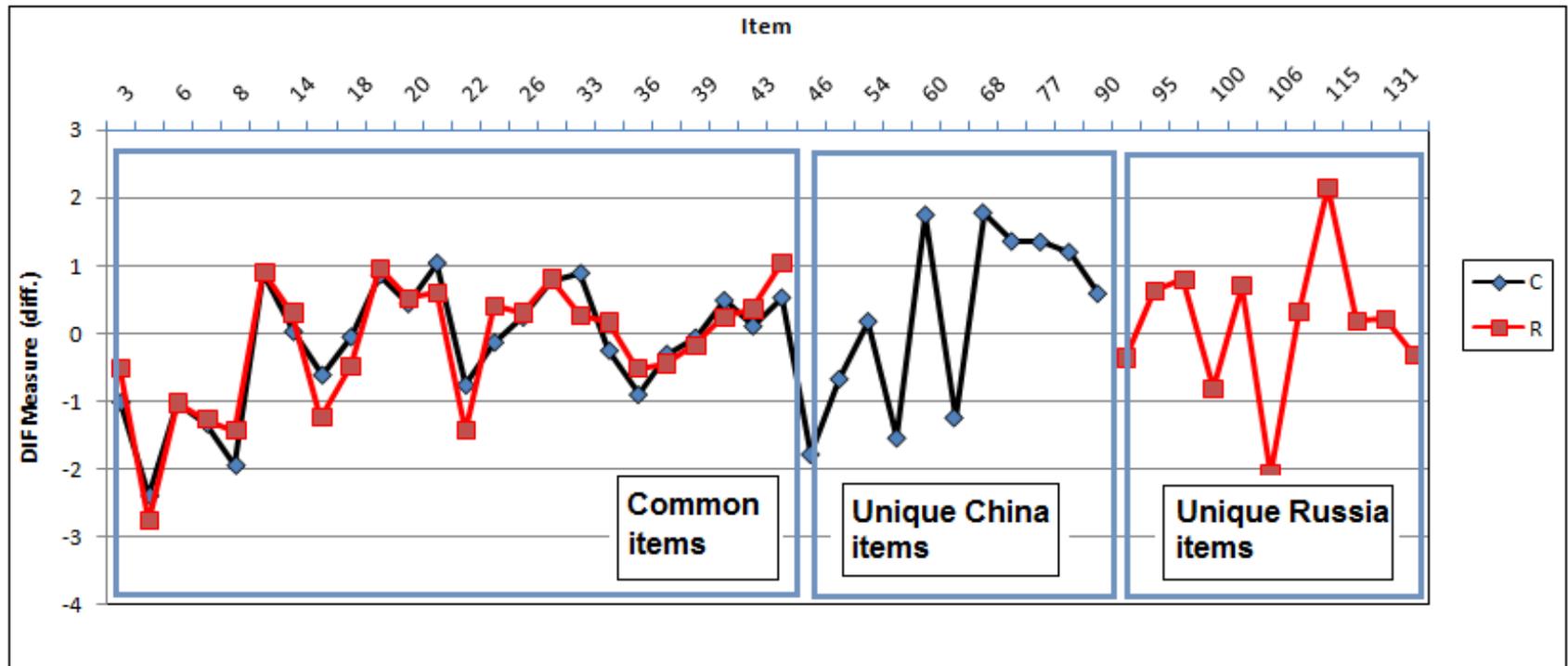
28 items are DIF free (no statistically significant DIF)

11 items with DIF (with statistically significant DIF): 5 items in favour of China, 6 items in favour of Russia

This suggests the test is not entirely fair for different countries

Our solution

- We used 28 DIF-free items for linking between the two countries.
- Altogether, the total data set has 50 items, from them 28 items are common for both countries, and 22 items are country-specific: 11 for China and 11 for Russia.

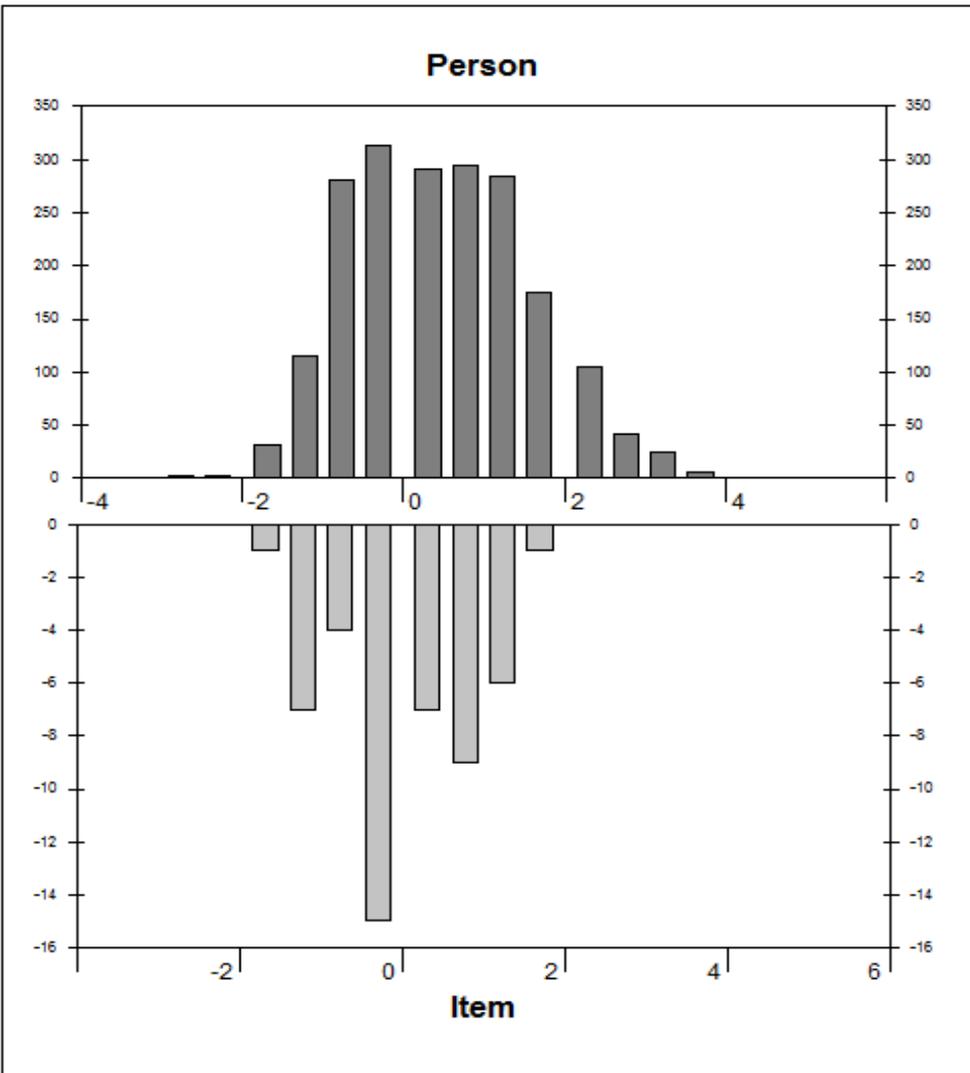


Psychometric analysis

Objective	Methodology
1. Choose “good” items 2. Make sure we measure “ability” as a unidimensional construct 3. Make sure items are fair across countries	1. Classical Test Theory + IRT 2. Test of dimensionality 3. DIF analysis
4. Look at the test as a whole: understand if the test is an appropriate measurement tool for the sample	4. Reliability study, variable map

Psychometric analysis.

Math test, grade 1



- The person reliability is 0.85 (classical reliability $\alpha = 0.83$).
- The person separation index is 2.39
- The test is essential unidimensional

Other tests showed similar results

Conclusion of psychometric analysis

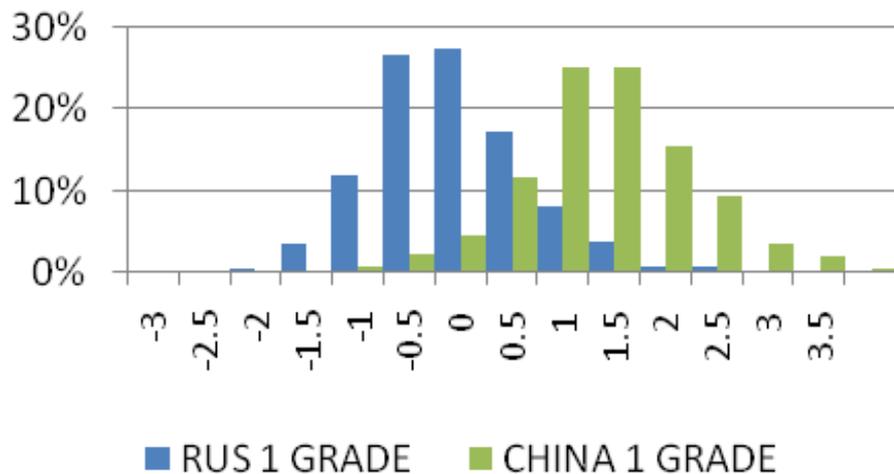
- The grade 1 and grade 3 mathematics and physics tests have good psychometric quality, are essentially unidimensional, reliable and valid instruments for measuring and comparing the mathematics and physics skills of grade 1 and grade 3 engineering students within and across the two countries.
- Although some test items demonstrated DIF across countries, a sufficient number of DIF-free items allowed us to construct a common scale for both countries.

Conclusion

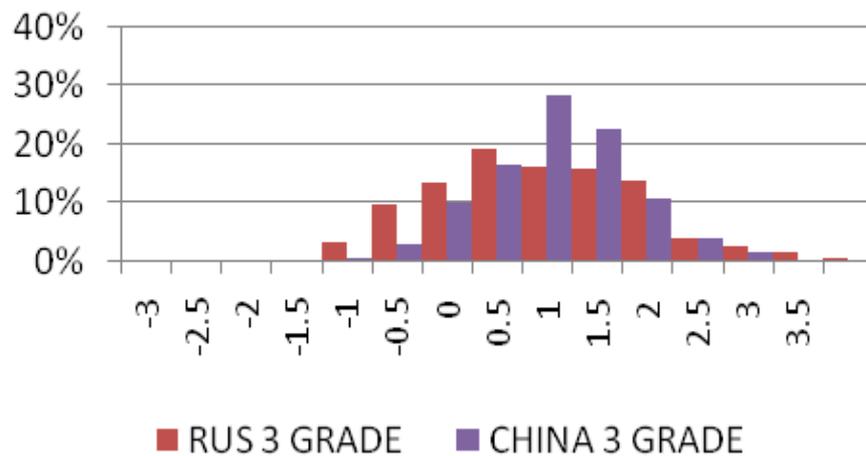
- We developed test instruments for assessing and comparing the mathematics and physics skill levels and gains of grade 1 and grade 3 engineering students in Russia and China.
- We showed that our test instruments are of high quality and can be used for international comparative research.

*The first results:
comparing
between
countries*

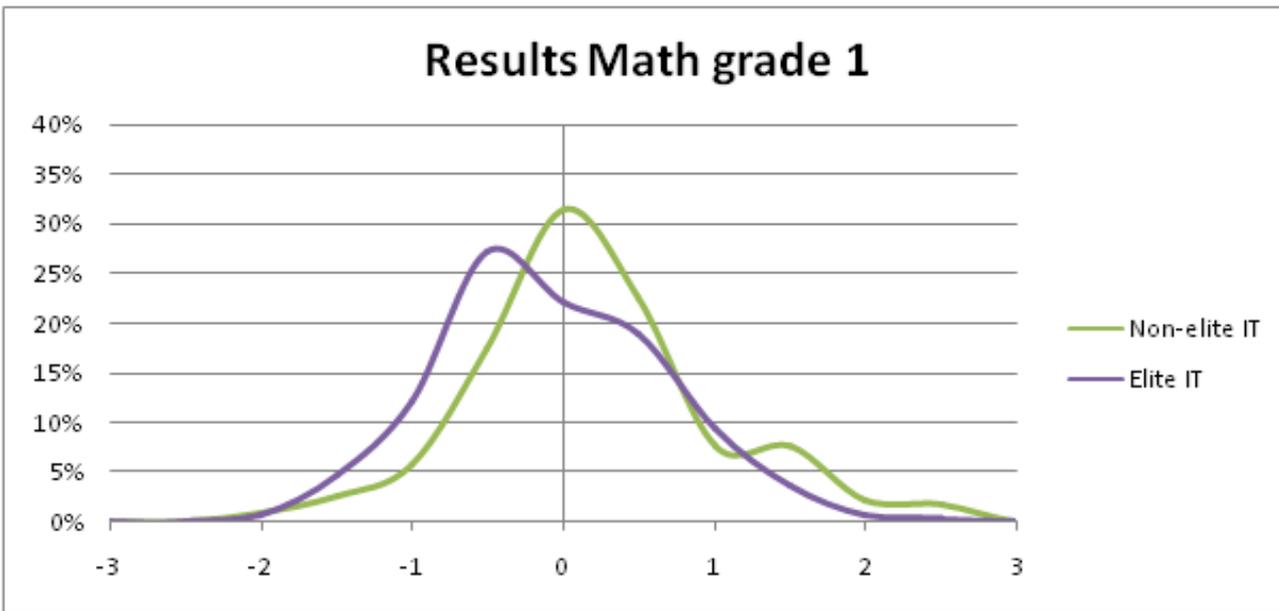
Math results



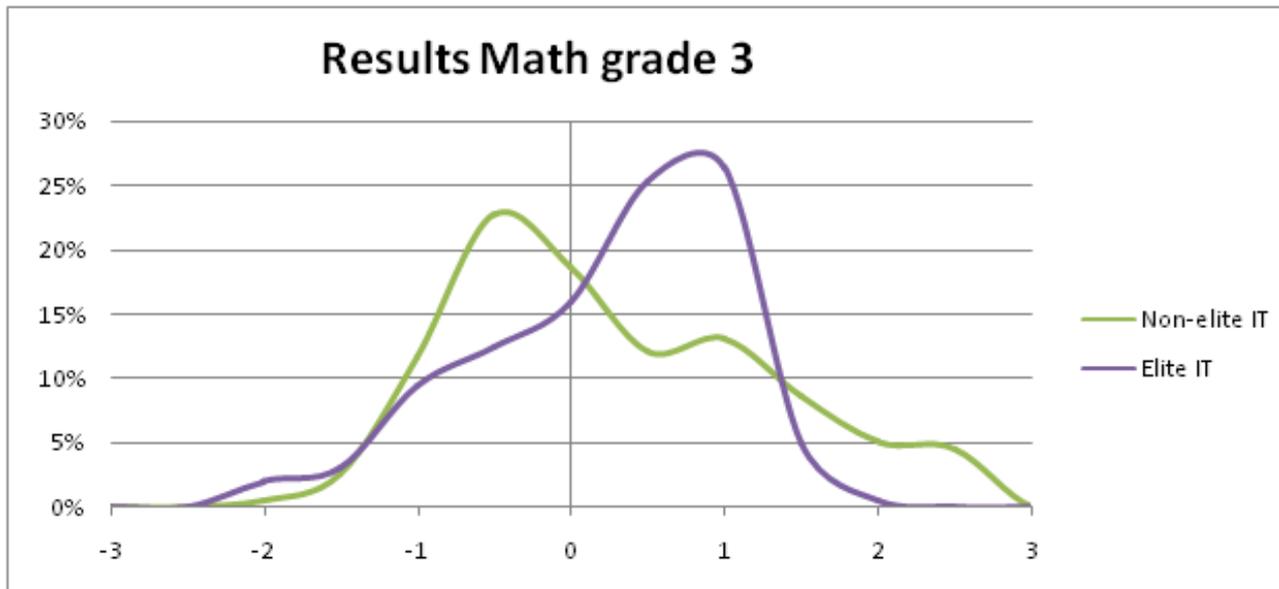
Math results



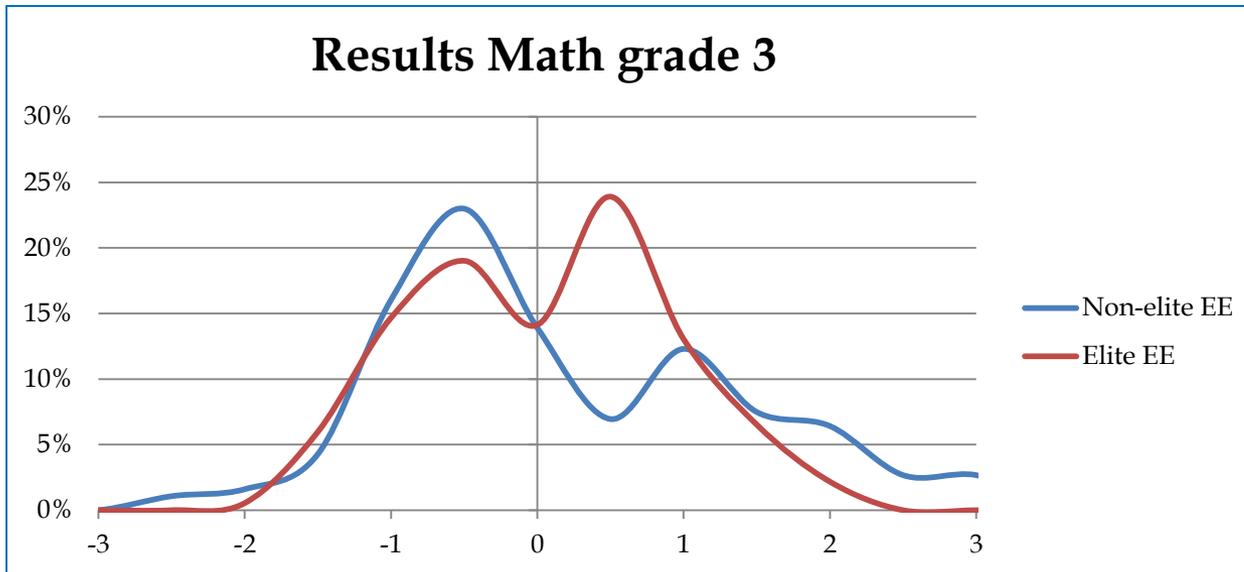
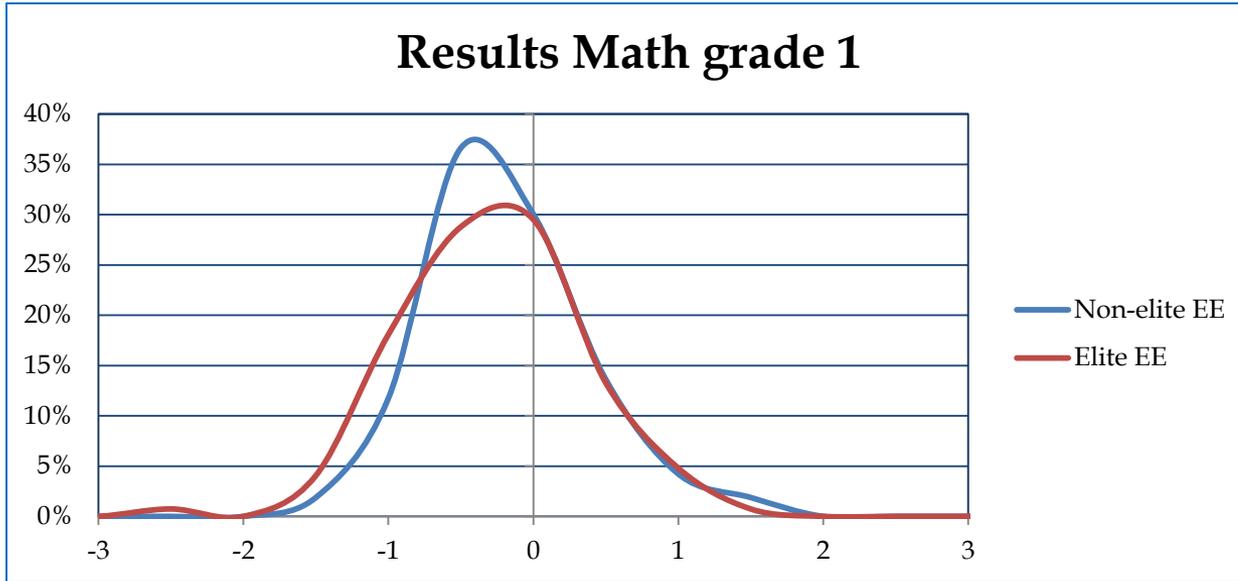
The first results (Russia)



Comparing elite and non-elite universities:
CS



The first results (Russia)



Comparing elite and non-elite universities:
EE

Thank you for your attention