

# The PISA Computer Based Assessment of Science: What did we learn?

Júlíus K. Björnsson



NÁMSMATSSTOFNUN

# Key findings

No differences for countries overall between science achievement on the paper and pencil test and science achievement on the computer-based test.

Boys outperform girls on the computer-based assessment of science in all countries.

The gender differences in performance cannot easily be linked to motivation, enjoyment or familiarity with computers.

# Unanswered questions

How much do we know about the effects on answering an item correctly by the response requirements?

Does it make a difference whether you tick a correct answer on a mark-sheet, write text in a paper-based environment or click a mouse, drag and drop or write text using a key board?

Do spelling and grammar play the same role in computer-based testing as in traditional paper and pencil?

# Unanswered questions

Does a move to an on-screen test affect male and female test-takers in the same way?

Can the same types of competencies be assessed via computer as by P&P methods?

Does the move to computer-based testing affects students in the same way across various subject areas such as reading, mathematics and science?

# Previous research

Previous research has in part indicated the psychometric equivalence of computer-based and paper and pencil tests

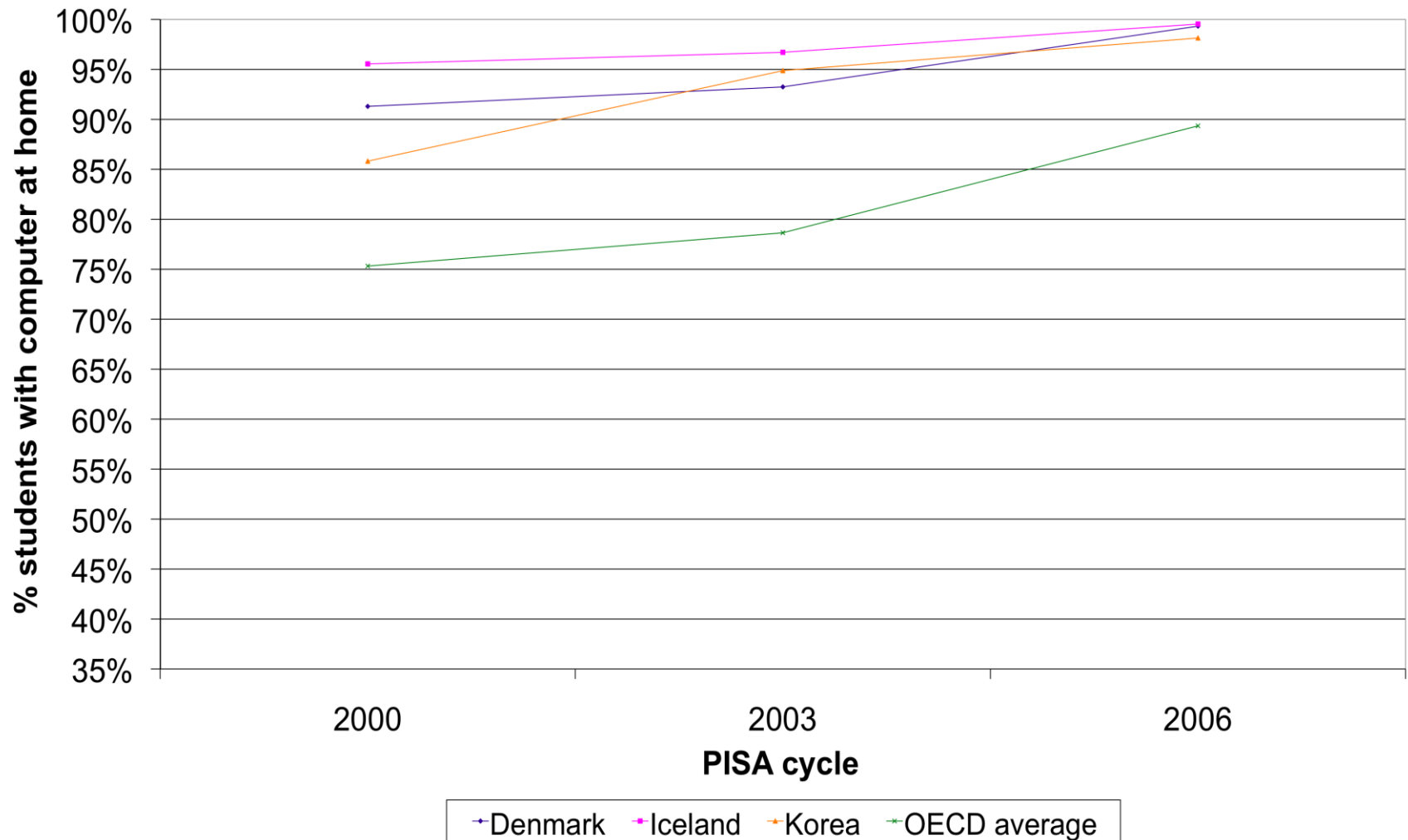
But other studies have highlighted a small gender difference in performance for tests that are performed on computer that was not present when the tests were performed via paper and pencil

AND things are changing faster and faster

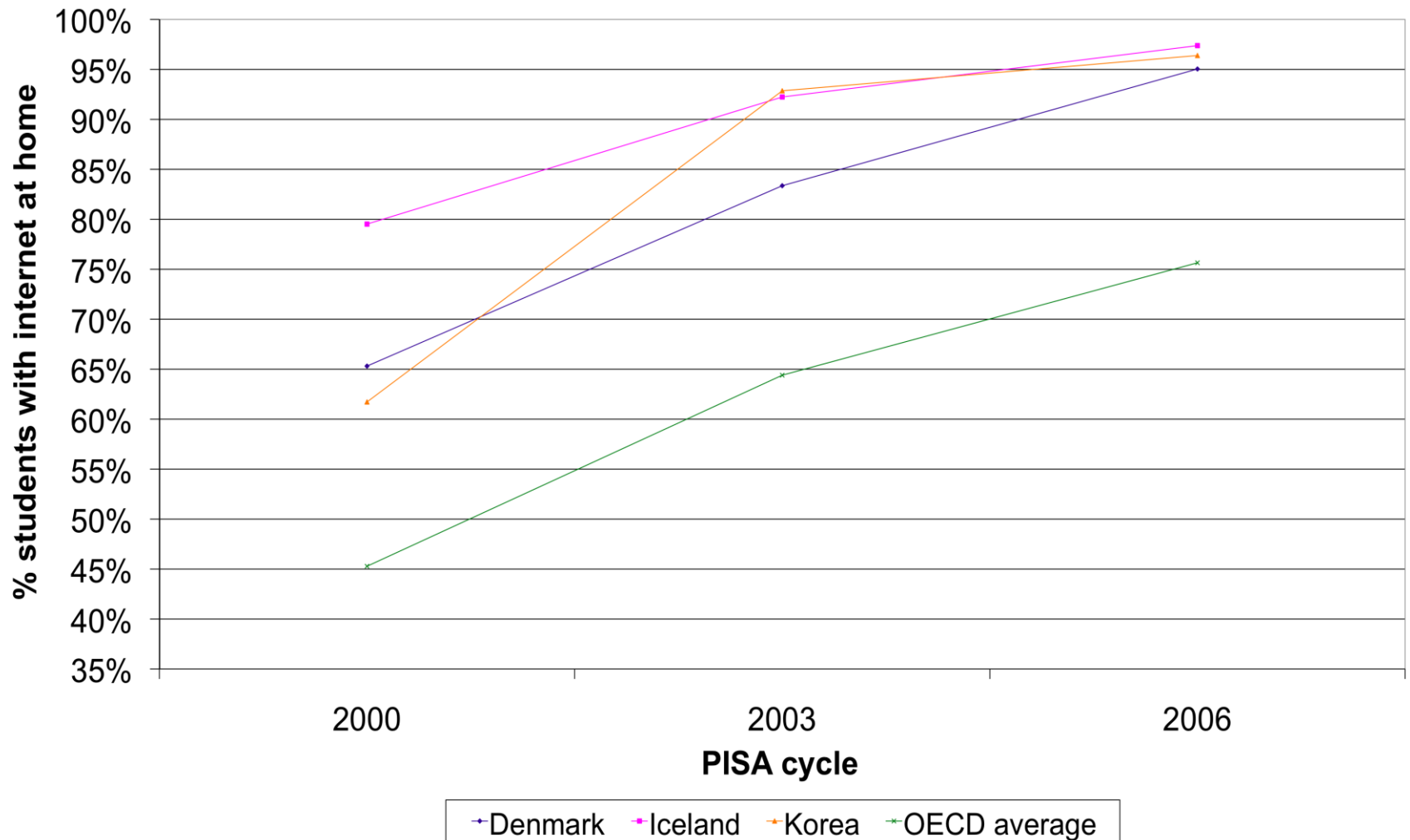
# Where do students learn?

	<i>How I learnt to use the internet</i>		<i>How I learnt to use computers</i>	
	<i>From myself</i>	<i>From friends</i>	<i>From myself</i>	<i>From friends</i>
Denmark	41.3%	18.4%	28.6%	18.5%
Iceland	47.6%	13.6%	35.7%	14.0%
Korea	52.1%	29.1%	35.7%	32.4%
OECD Average*	33.9%	16.7%	26.6%	14.2%

# Computers at home

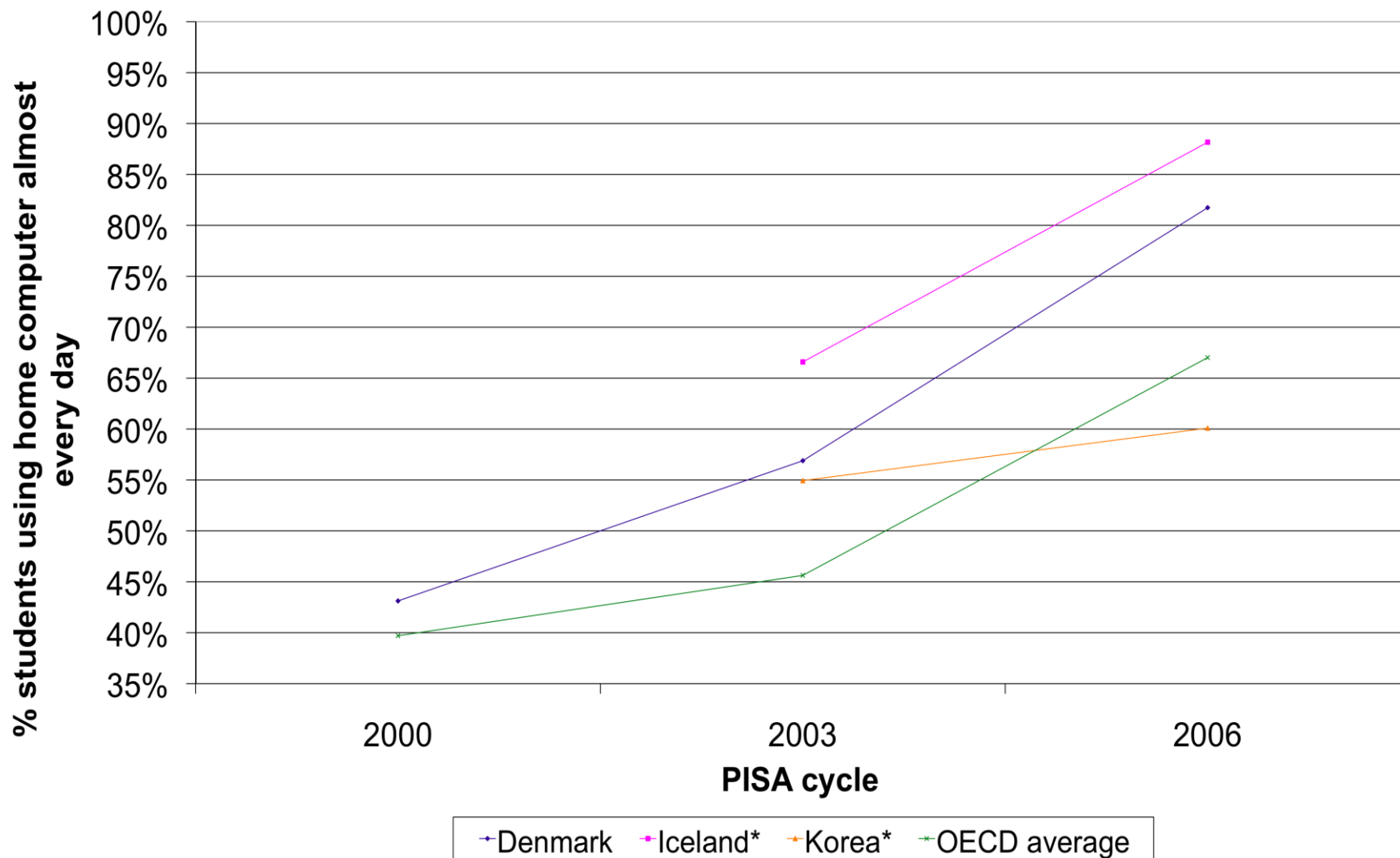


# Internet at home

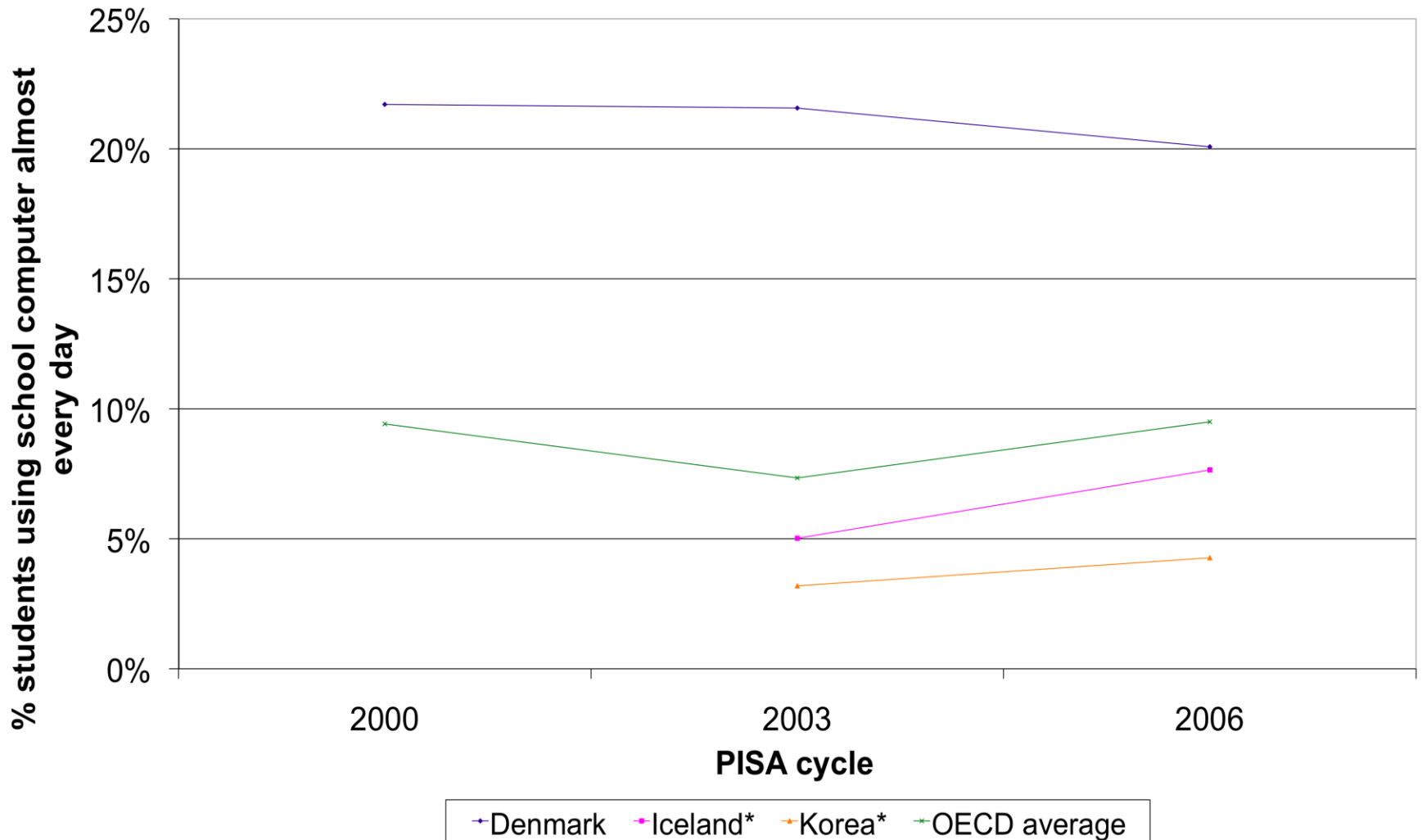




# Computer use at home every day



# Computer use in school



# CBAS aims

Administer questions that would be difficult to deliver in a paper and pencil test.

Reduce the amount of reading required.

Compare paper and pencil and electronic assessment of Science.

Try out computerized test administration in an international setting.

# Implementation

Field trial in 2005, 13 participants.

Main study in 2006. Only 3 participants, Denmark, Iceland and Korea.

A subsample of 100 schools was selected to participate in CBAS from the main PISA 2006 school sample test in Iceland

Clusters of 5 to 45 PISA-eligible students were sampled from the PISA student sample from each school.

The Icelandic sample for CBAS is representative of the population of 15 year old students in the country.

# Implementation

In Denmark and Korea, approximately half of the PISA sampled schools were sampled for CBAS.

Test administrators travelled between schools, with 6-8 laptops each, setting up a standardized testing environment in each one.

All countries used the same type of laptop, the same software and procedures.

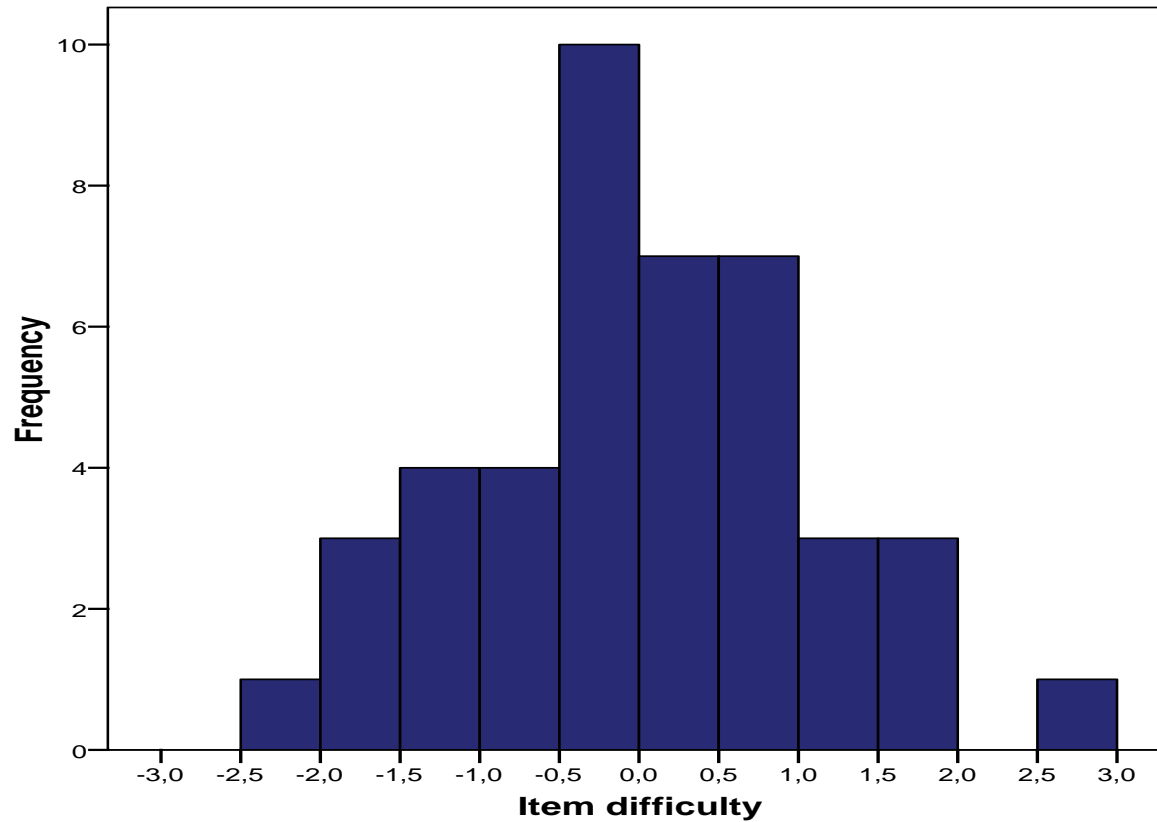
# The test

45 items with multimedia extensions (animations using flash software, video footage or photos) were presented to students.

All item designs were either multiple choice or complex multiple choice involving, for example, a number of Yes/No responses for the answers offered.

A small number of complex multiple choice items asked the students to place items in a specific order or position in a given diagram.

# Item difficulty



Mean =0,01027  
Std. Dev. =1,11964  
N =43

# Example-low interactivity

## Question 19: Assembly Line

Robots are used to make cars on assembly lines.



How does this robot technology affect human society? Answer "Yes" or "No" for each statement.

Consumer goods are made more cheaply.

Yes

No

Workers avoid exposure to some hazardous materials.

Yes

No

The type of jobs workers do remains the same.

Yes

No

Robots make fewer mistakes than humans.

Yes

No

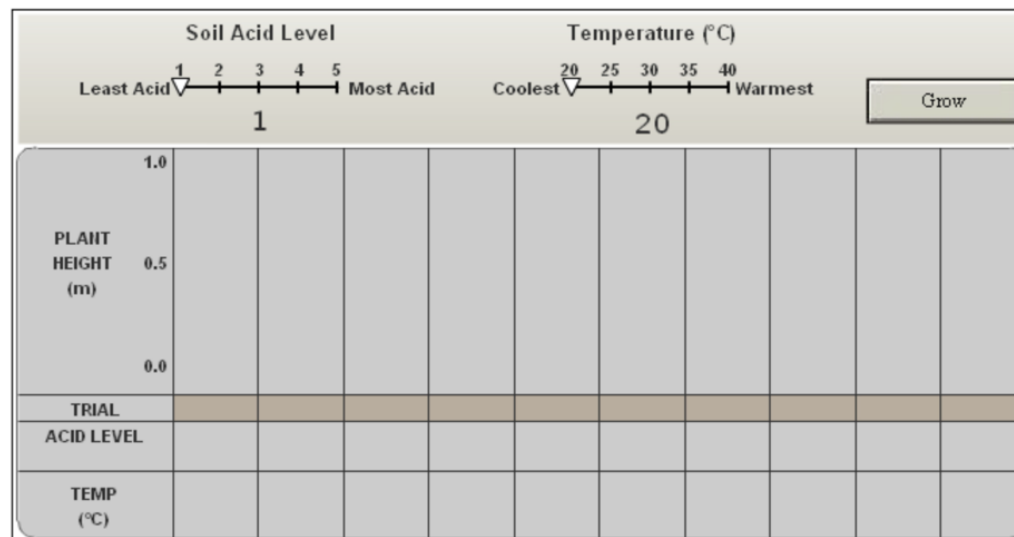


# Example-high interactivity

## Question 13: Plant Growth

The height of wheat plants is affected by temperature and the soil acid level. At what soil acid level and temperature does this new variety of wheat grow tallest?

Use the sliders to set your variables (soil acid level and temperature) and then press the "Grow" button. The plant will grow, and the average height of the wheat after 1 month will be shown. You may choose up to ten different combinations of variables for your experiment.



Select the combination of soil acid level and temperature that produces the tallest wheat.

Soil Acid Level:

- 1     2     3     4     5

Temperature (°C):

- 20°     25°     30°     35°     40°

# Additionally

Each student answered online a short questionnaire about the experience.

Each student assessed the difficulty of the test.


*How much effort did you invest?*

Please try to imagine an actual situation (at school or in some other context) that is highly important to you personally, so that you would try your very best and put in as much effort as you could to do well.

In this situation you would mark the highest value on the "effort thermometer", as shown below:

Compared to the situation you have just imagined, how much effort did you put into doing this test?

How much effort would you have invested if your marks from the test were going to be counted in your school marks?

	<input type="checkbox"/> 10	<input type="checkbox"/> 10	<input type="checkbox"/> 10
	<input checked="" type="checkbox"/> 9	<input type="checkbox"/> 9	<input type="checkbox"/> 9
	<input type="checkbox"/> 8	<input type="checkbox"/> 8	<input type="checkbox"/> 8
	<input type="checkbox"/> 7	<input type="checkbox"/> 7	<input type="checkbox"/> 7
	<input type="checkbox"/> 6	<input type="checkbox"/> 6	<input type="checkbox"/> 6
	<input type="checkbox"/> 5	<input type="checkbox"/> 5	<input type="checkbox"/> 5
	<input type="checkbox"/> 4	<input type="checkbox"/> 4	<input type="checkbox"/> 4
	<input type="checkbox"/> 3	<input type="checkbox"/> 3	<input type="checkbox"/> 3
	<input type="checkbox"/> 2	<input type="checkbox"/> 2	<input type="checkbox"/> 2
	<input type="checkbox"/> 1	<input type="checkbox"/> 1	<input type="checkbox"/> 1

# Students' achievement in Science

Overall achievement within countries did not change from one test modality to the next. (Yet, there was a tendency for Denmark's performance to decrease on the computer-based test).

Korean students outperformed Danish and Icelandic students in the computer-based test just as they did in the paper and pencil test.

Boys' performance increased in Iceland and Korea in the computer-based test while girls' performance decreased.

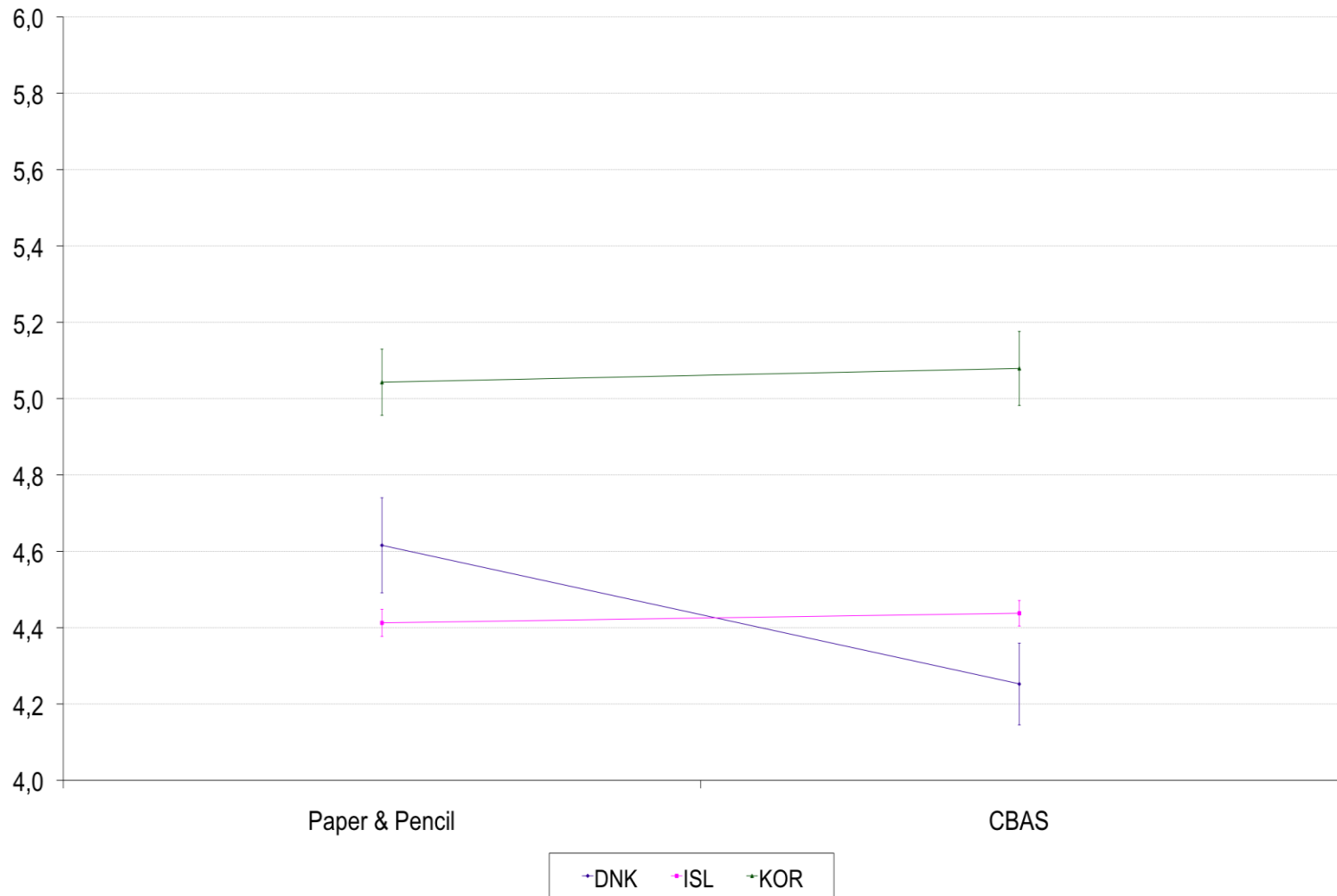
# Students' achievement in Science

Boys outperformed girls on the computer-based test in all three countries

Girls outperformed boys on the paper and pencil test of science in Iceland whereas there was a gender difference in favour of boys in the paper and pencil results for Denmark.

The association between reading literacy and achievement on the science test was less strong for the computer-based items than for the paper and pencil items.

# Paper and pencil compared to CBAS



## Correlations between P&P science scores and CBAS

<i>Correlations</i>	<i>Girls</i>	<i>Boys</i>
Denmark	0.89	0.91
Iceland	0.78	0.80
Korea	0.88	0.90

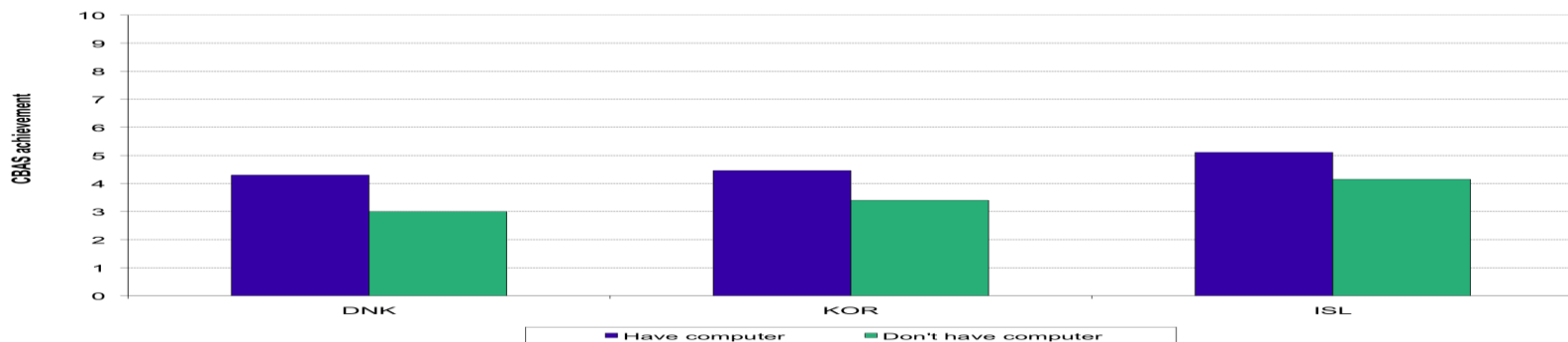
The correlations show that girls' CBAS performance is slightly less strongly associated with their performance on the P&P test of science than for boys, indicating that **the impact of changing the test method is not the same for girls as it is for boys.**

## Home ICT access and achievement in CBAS

The vast majority of students have computers at home and internet connections

Not having a computer at home is associated with poorer performance on the CBAS test

	<i>Have computer at home</i>	<i>Have educational software at home</i>	<i>Have internet access at home</i>	<i>Number of computers in home</i>
Denmark	98%	58%	96%	51% have three or more
Iceland	98%	69%	98%	45% have three or more
Korea	97%	62%	96%	73% have one



## *ICT frequency of use and confidence across genders*

Overall, boys score higher on the frequency of use scales than girls.

Perhaps as a result of this, they also score higher on the confidence in using ICT scales, particularly in Iceland.

Girls tend to use the internet more for social networking activities such as chatting and email and their confidence is consequently higher for these activities.

Boys tend to browse the internet, play games and download software a lot more than girls and they perform advanced computer activities more frequently.

Overall, boys have far greater confidence in most ICT activities which is similar to ICT PISA 2003 results (OECD, 2005b).



# Frequency and Confidence

Scores from two ICT scales relating to ICT use and from the two scales relating to confidence were added to get an overall ICT Confidence score and an overall Frequency of Use score. Scores on these scales were dichotomised into two equally sized groups: ICT frequent and infrequent users and ICT confident and unconfident.

<i>Country</i>	<i>Girls</i>		<i>Boys</i>	
	<i>ICT infrequent users</i>	<i>ICT frequent users</i>	<i>ICT infrequent users</i>	<i>ICT frequent users</i>
Denmark	62.7%	37.3%	36.8%	63.2%
Iceland	62.5%	37.5%	37.4%	62.6%
Korea	54.3%	45.7%	46.6%	53.4%
<i>TOTAL</i>	<i>59.8%</i>	<i>40.2%</i>	<i>40.3%</i>	<i>59.7%</i>

# Frequency and Confidence

What is notable in these patterns is that although over 60% of Icelandic girls report that they are frequent users of the Internet and programs, only 33% report that they feel confident about performing ICT tasks.

	<i>Girls</i>		<i>Boys</i>	
	<i>ICT Low confidence</i>	<i>ICT high confidence</i>	<i>ICT low confidence</i>	<i>ICT high confidence</i>
Denmark	64.1%	35.9%	35.8%	64.2%
Iceland	66.6%	33.4%	36.3%	63.7%
Korea	49.3%	50.7%	50.4%	49.6%
<i>TOTAL</i>	<i>60.0%</i>	<i>40.0%</i>	<i>40.8%</i>	<i>59.2%</i>

Moderate relationships to performance on both tests

# CBAS questionnaire results

Boys' responses to the questionnaire items are more polarised than girls'. They usually strongly disagree and strongly agree with the statements more than girls, who tend towards the middle categories.

Motivation for the computer-based test was higher than for the paper and pencil test across all countries.

Students enjoyed the computer-based test more than the paper and pencil test.

Most students prefer to do a computer-based test than a paper and pencil test.

Most students reported that they put the same amount of effort into both tests.

Test enjoyment and motivation seem to have little relationship with achievement.

Test preference and relative effort report show no association with test performance.

## Features of the computer-based items and performance

The higher reading load on paper and pencil science items appears to disadvantage boys.

The gender difference is removed when we look at performance on paper and pencil items that are of low reading load.

Boys do better than girls on more difficult paper and pencil and computer-based items.

Domain coverage across test modalities is similar but boys outperform girls in all domains on the computer-based items.

There were more computer-based items that were easier for boys than there were items that were easier for girls.

# Summary

## **PISA Science vs. CBAS difference**

In all participating countries, there was no evidence to suggest that overall group performance was affected by the method of test presentation (computer-based or paper and pencil). However, there was a slight trend in Denmark for the scores on the computer-based test to drop.

## **Country differences**

Korea scored higher than both Iceland and Denmark on the paper and pencil assessment of science as well as on the computer-based assessment of science.

# Summary

## **Gender differences**

On the paper and pencil test of science in Iceland, girls outperformed the boys and in Denmark boys outperformed the girls. However, when the test was administered via computer, boys in all three countries outperformed their female counterparts.

## **Explanations**

In investigating the features of the computer-based test we found that boys found the shorter reading load items comparatively easier than girls. It appears that the paper and pencil science test disadvantages boys due to the length of the items. To the same extent, computer presentation seems to disadvantage girls, but this is more difficult to explain.

# Summary

## **Test motivation and enjoyment**

Students on the whole enjoyed the computer-based test more than the paper and pencil and were more motivated to perform another computer-based test than another paper and pencil test.

## **Test motivation, enjoyment and achievement**

Although we found a relationship between the effort reported by students and achievement on the paper and pencil test, this was not present for the computer-based test.

# Conclusion

In conclusion, any changes to methods of assessment should be made with caution and preferably after an initial analysis comparing achievement on a paper and pencil test with achievement on a computer-based test of the same paper and pencil items and achievement on computer-based items in the same domain.