PISA Computer-Based Assessment of Student Skills in Science

Programme for International Student Assessment

OECD

PISA
Computer-Based Assessment of Student Skills in Science

Programme for **I**nternational **S**tudent **A**ssessment

OECD

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

ISBN 978-92-64-08202-1 (print)
ISBN 978-92-64-08203-8 (PDF)

Photo credits: Cover © Left: Ariel Skelley/Blend Images/Getty Images
© Center: Spohn Matthieu/PhotoAlto Agency RF Collections/Getty Images
© Right: Geostock/Photodisc/Getty Images

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

Foreword

The capacity of countries – both the world's most advanced economies as well those experiencing rapid development – to compete in the global knowledge economy increasingly depends on whether they can meet a fast-growing demand for high-level skills. This hinges on significant improvements in the quality of schooling outcomes and a more equitable distribution in learning opportunities.

The OECD's Programme for International Student Assessment (PISA) makes it possible to regularly and directly compare the quality of educational outcomes across education systems. PISA assesses some of the key competencies contributing to the success of individuals and societies on a regular basis and within an international framework. The results provide a basis for policy dialogue and collaboration in defining and implementing educational goals in innovative ways that reflect judgements about the skills relevant to adult life. Across the globe – whether in Canada in North America, Finland in Europe or Japan and Korea in Asia – countries in PISA show that excellence in education is an attainable goal, and often at reasonable cost. They also show that the challenge of achieving a high and socially equitable distribution of learning outcomes can be successfully addressed and that excellence can be achieved consistently throughout the education systems, with very few students and schools left behind.

In the same way as the nature of the competencies that contribute to the success of individuals continues to change, the assessments such as PISA also need to adapt and advance. There are at least three major groups of challenges ahead for the development of future PISA assessment instruments, all of which relate to technology.

The first, and perhaps most important group of challenges, relates to embracing a wider range of competencies than are currently amenable to a large-scale assessment. State-of-the art skills in fields like reading, mathematics and science that are currently assessed by PISA will always be important; as much as modern educators value "learning to learn" skills, people only learn by learning something. However, the more complex the world becomes, the more success will hinge on the capacity of people to collaborate and orchestrate complex processes. Innovation itself is rarely the product of individuals working in isolation. A challenge for future assessments will be not only to value individually demonstrated competencies but to also introduce interpersonal competencies. Likewise, the conventional approach to problems has been to break them down into manageable bits and pieces and then teach students the techniques to solve the pieces and assess the extent to which they have successfully mastered the techniques. Modern societies, however, increasingly create value by synthesising disparate fields of knowledge. This requires curiosity and the ability to make connections between ideas that previously seemed unrelated, which in turn requires being familiar with and receptive to knowledge in other fields than one's own. Assessments that are limited to the reproduction of isolated subject-matter knowledge in multiple-choice format will not be able to reflect that. Furthermore, the more content people can search and access, the more important becomes the capacity to make sense out of this growing and unstructured body of knowledge, to understand how students question or seek to improve the accepted knowledge and to know how students spend their time. In the past, teachers could tell students to use an encyclopaedia when they needed some information, and students could rely on what they found there to be true. Today, literacy is about managing non-linear information structures, dealing with ambiguity and managing and resolving conflicting pieces of information found somewhere on the web. The challenge for future assessments will be to integrate this technology-rich notion of literacy. Last but not least, in the past some used to divide the world into specialists and generalists. Specialists generally have deep skills and a narrow scope, giving them expertise that is recognised by peers but not

valued outside their domain. On the other hand, generalists have a broad scope but their skills may lack in depth. Success today, however, requires one to be versatile, apply a depth of skills to a progressively widening scope of situations and experiences, gain new competencies, build relationships and assume new roles. It requires being capable not only of constantly adapting but also constantly learning and growing. The challenge for future assessments will therefore be to move from providing static tasks to ones that are dynamic and interact with students.

A second group of challenges relates to the delivery of assessments. Paper-and-pencil assessments are inadequate to deliver authentic tasks. Future assessments will need to be electronically delivered, interactive and dynamic. Computer-assisted scoring will also help to make the assessment process more efficient and narrow the time lag between collecting the data and making results available to feed into educational improvement.

A third group of challenges relates to the relevance of the assessment tasks. One-size-fits-all assessments that use a single set of instruments to assess all students inevitably lead to significant frustration for both high and low performing students, who are then confronted with large numbers of tasks that are too easy or too difficult. Computer-based assessments have the potential to interactively adapt to the skill potential of students and thus provide a much more effective and engaging assessment experience.

Addressing these challenges is a huge task, particularly in an international context. However, PISA has begun to make many important advances in this area. This report documents the first step into the direction of dynamic, interactive and adaptive electronically delivered PISA assessments that was pioneered by Denmark, Iceland and Korea. While the results from this first computer-based assessment of science highlight numerous challenges, not all of which are fully understood yet, they have encouraged countries to take the work further. In 2009, 17 countries took part in a much larger exercise to assess the reading of digital texts and for 2012, a computer-based assessment of dynamic problem-solving skills is planned. The OECD expresses its deep gratitude to the three countries who have pioneered this work and laid the foundation for further work in this area.

The report is the product of a collaborative effort between Denmark, Iceland and Korea, the experts and institutions working within the framework of the PISA Consortium, and the OECD.

The report was drafted at the Educational Testing Institute in Reykjavik Iceland by Pippa McKelvie, with advice as well as analytical and editorial support from Almar M. Halldórsson and Júlíus K. Björnsson.

The development of the computer-based assessment of science competencies was steered by the PISA Governing Board, which is chaired by Lorna Bertrand (United Kingdom).

The report is published on the responsibility of the Secretary-General of the OECD.

Lorna Bertrand
Chair of the PISA Governing Board

Barbara Ischinger
Director for Education, OECD

Table of Contents

FOREWORD 3

CHAPTER 1 **INTRODUCTION** 11

Key findings from the PISA 2006 Computer-based Assessment of Science 12
What is PISA? 12
What is a Computer-based Assessment of Science? 12
Objectives of the PISA 2006 Computer-based Assessment of Science (CBAS) 17
Notes 18

CHAPTER 2 **ADMINISTRATION OF THE CBAS TEST** 19

Field Trial 20
School and student sampling 20
- Sample weighting 20
- Gender distribution in the weighted sample 20

Administration of test sessions 21
- Technical specifications 21
- Cognitive item information 22
  - *Item order and missing responses* 22
  - *Item difficulty* 23
  - *Item difficulty across genders* 23
  - *Science framework classification* 23
  - *Interactivity* 24
  - *Reading load* 25
- CBAS questionnaire items 27

Notes 28

CHAPTER 3 **STUDENTS' ACHIEVEMENT IN SCIENCE** 29

Achievement across test modalities 30
- Scaling information 30
- Performance across countries and test modalities 30
- Strength of relationship between paper-and-pencil scores and CBAS scores 31
- Gender differences in student performance across test modalities 32

Notes 34

CHAPTER 4 **USE OF INFORMATION COMMUNICATION TECHNOLOGIES (ICT) ACROSS GENDERS AND IMPACT ON ACHIEVEMENT** 35

Home ICT access and achievement in CBAS 36
ICT familiarity items and scales 37
- ICT Internet/entertainment use 37
- ICT program/software use 37

- Self-confidence in ICT Internet tasks 37
- Self-confidence in ICT high-level tasks 37

ICT frequency of use and confidence across genders 39

- Frequency of performing specific ICT activities 39
- Confidence in performing specific ICT activities 40
- Comparison of male and female confidence and frequency of use 41

Association between ICT familiarity, confidence and CBAS achievement 42

- Impact of frequency of ICT use on science achievement 42
- Impact of ICT confidence on science achievement 44
- ICT familiarity and CBAS achievement per country 48

Notes 51

CHAPTER 5 **CBAS QUESTIONNAIRE RESULTS** 53

Test enjoyment across test modalities and genders 55

Test motivation across test modalities and genders 57

Relative effort applied to tests across genders 59

- Preference for the computer-based or paper-and-pencil test 59
- Impacts of motivation and enjoyment on achievement 61

Impacts of test preference and relative effort on achievement 65

Students' reported effort on PISA and CBAS: What the effort thermometer shows 68

Relationship between reported effort and achievement 70

Notes 74

CHAPTER 6 **FEATURES OF THE COMPUTER-BASED ITEMS AND PERFORMANCE** 75

Missing responses and gender 77

Reading load 78

- Reading load and CBAS achievement across genders 78
- Reading load and science achievement for Iceland 79
- Comparison of very low reading load paper-and-pencil items and CBAS items 80

Interactivity of items 80

Gender differences in percentage correct per item 82

- Gender differences in performance and item difficulty 86

Achievement across sub domains of science 87

- Paper-and-pencil science performance across domains in Iceland 88

Notes 89

CHAPTER 7 **SUMMARY** 91

Explaining the gender difference 92

- Was this a "test for males"? 92

REFERENCES 94

ANNEX A **SAMPLING INFORMATION** 95

ANNEX B **FREQUENCY OF PERFORMING ICT ACTIVITIES BY COUNTRY AND GENDER** 97

FIGURES

Figure 1. Proportion of students possessing a computer in the home over PISA cycles....................... 15
Figure 2. Average percentage of students with Internet connection at home across PISA cycles.......... 16
Figure 3. Average percentage of students using a computer at home almost every day 16
Figure 4. Average percentage of students using a computer at school almost every day...................... 17
Figure 5. Distribution of item difficulty for final CBAS items... 22
Figure 6. Sample unit: Assembly Line.. 24
Figure 7. Sample unit of highly interactive item: Plant Growth 25
Figure 8. Sample item showing high reading load item: Echolocation 26
Figure 9. Sample item showing low reading load item: Bean Leaves 26
Figure 10. PISA Effort Thermometer.. 27
Figure 11. Achievement in paper-and-pencil test of science compared to CBAS.................................. 30
Figure 12. Average CBAS score and average paper-and-pencil science scores in Denmark (N=1254)... 31
Figure 13. Average CBAS score and average paper-and-pencil science scores in Iceland (N=3566)...... 32
Figure 14. Average CBAS score and average paper-and-pencil science scores in Korea (N=2650) 32
Figure 15. Male achievement advantage across tests and countries (positive values show males outperforming females)... 33
Figure 16. Performance on CBAS and possession of computer in the home per country....................... 36
Figure 17. Students' use of the Internet and average performance in CBAS by quarter of the indices 42
Figure 18. Students' use of the Internet and average performance in paper-and-pencil science by quarter of the indices 43
Figure 19. Students' use of programs and educational software and average performance in CBAS by quarter of the indices 44
Figure 20. Students' use of programs and educational software and average performance in paper-and-pencil science by quarter of the indices 45
Figure 21. Internet confidence and CBAS achievement across countries and genders.......................... 45
Figure 22. Internet confidence and paper-and-pencil science achievement across countries and genders.. 46
Figure 23. Advanced ICT activities confidence and CBAS achievement across countries and genders .. 46
Figure 24. Advanced ICT activities confidence and paper-and-pencil science achievement across countries and genders.. 47
Figure 25. Impact of ICT familiarity for male and female students on CBAS and paper-and-pencil scores in Denmark... 47
Figure 26. Impact of ICT familiarity for male and female students on CBAS and paper-and-pencil scores in Iceland.. 48

Figure 27. Impact of ICT familiarity for male and female students on CBAS and paper-and-pencil scores in Korea 49

Figure 28. Danish students' endorsement of the statement "I found the CBAS test enjoyable" (top) and endorsement of the statement "I found the paper-and-pencil test enjoyable" (bottom) 55

Figure 29. Icelandic students' endorsement of the statement "I found the CBAS test enjoyable" (top) and endorsement of the statement "I found the paper-and-pencil test enjoyable" (bottom) 56

Figure 30. Korean students' endorsement of the statement "I found the CBAS test enjoyable" (top) and endorsement of the statement "I found the paper-and-pencil test enjoyable" (bottom) 56

Figure 31. Danish students' endorsement of the statement "I would do another computer-based test for fun" (top) and endorsement of "I would do another paper-and-pencil test for fun" (bottom) 57

Figure 32. Icelandic students' endorsement of the statement "I would do another computer-based test for fun" (top) and endorsement of "I would do another paper-and-pencil test for fun" (bottom) 58

Figure 33. Korean students' endorsement of the statement "I would do another computer-based test for fun" (top) and endorsement of "I would do another paper-and-pencil test for fun" (bottom) 58

Figure 34. Male and female students' relative effort across assessments 60

Figure 35. Students' preference for the computer-based or paper-and-pencil test 61

Figure 36. Motivation for CBAS test and achievement 62

Figure 37. Enjoyment of CBAS test and achievement 62

Figure 38. Motivation for paper-and-pencil test and paper-and-pencil science achievement 63

Figure 39. Enjoyment of paper-and-pencil test and paper-and-pencil science achievement 64

Figure 40. Average CBAS and paper-and-pencil science achievement and test preference for Danish females (top) and males (bottom) 65

Figure 41. Average CBAS and paper-and-pencil science achievement and test preference for Icelandic females (top) and males (bottom) 66

Figure 42. Average CBAS and paper-and-pencil science achievement and test preference for Korean females (top) and males (bottom) 66

Figure 43. Average CBAS and paper-and-pencil science achievement and relative effort on tests for Danish females (top) and males (bottom) 67

Figure 44. Average CBAS and paper-and-pencil science achievement and relative effort on tests for Icelandic females (top) and males (bottom) 67

Figure 45. Average CBAS and paper-and-pencil science achievement and relative effort on tests for Korean (females (top) and males (bottom) 68

Figure 46. Effort differential for CBAS and PISA across countries 69

Figure 47. How much harder students would have tried if the results of CBAS and the paper-and-pencil test counted for school marks 70

Figure 48. CBAS reported effort and achievement across countries 71

Figure 49. PISA reported effort and PISA science achievement 71

Figure 50. Distribution of effort thermometer scores in PISA paper-and-pencil test 73

Figure 51. Distribution of effort thermometer scores in CBAS test 73

Figure 52. Relationship between item order and item difficulty 76

Figure 53. Relationship between item order and amount of missing responses 77

Figure 54. Average percentage difference in achievement between males and females on high and low reading load CBAS items 79

Figure 55. Average percentage difference in achievement between males and females according to reading load across test modalities in Iceland 79

Figure 56. Gender difference in percentage correct depending on reading load of items in Iceland 80

Figure 57. Percentage correct on high and low interactivity CBAS items 81

Figure 58. Average performance advantage for males across countries according to the interactivity of CBAS items. 82

Figure 59. Sample unit: Ping Pong Ball Q1 83

Figure 60. Sample unit: Ping Pong Ball Q2 83

Figure 61. Sample item showing bias towards male: Bicycle Q1 84

Figure 62. Sample item showing bias towards male: Designed for Flight Q1 84

Figure 63. Sample item showing bias towards female: Nuclear Power Plant 85

Figure 64. Sample item showing bias towards female: Washing Glasses Q3 85

Figure 65. Percentage correct advantage for males on high and low difficulty CBAS items. 86

Figure 66. Gender difference in performance on harder and easier paper-and-pencil items 87

Figure 67. Gender difference in performance across domains of science in CBAS 87

Figure 68. Average advantage in performance across domains for CBAS and paper-and-pencil science items in Iceland 88

TABLES

Table 1. Sources of learning about computers and the Internet (PISA 2003) 15

Table 2. Proportion of females and males in the sample analysed in this report 21

Table 3. Minimum hardware specifications for student laptops used for CBAS 21

Table 4. Distribution of paper-and-pencil science and CBAS items across domains of science 24

Table 5. Correlations for male and female students' performance in CBAS and paper-and-pencil tests across countries 31

Table 6. Achievement in paper-and-pencil test of science compared to CBAS 33

Table 7. Correlations between paper-and-pencil science scores and CBAS scores across genders and countries 33

Table 8. Proportion of students in each country with a computer, educational software and Internet at home 36

Table 9. Model fit for CFA with ICT familiarity items 38

Table 10. Estimated latent correlations for ICT familiarity scales 38

Table 11. Scale reliabilities for ICT familiarity scales 39

Table 12. Proportion of females and males reporting frequent or less frequent use of ICT across countries 41

Table 13. Proportion of females and males reporting high or low confidence in performing ICT tasks across countries 41

Table 14. Score point difference on science scale associated with one unit on the ESCS (gradient) 48

Table 15. Effects of ICT familiarity on performance for male and female students across countries 49

Table 16. Correlations between ICT familiarity science scores on the CBAS and paper and pencil tests. 50

Table 17. Male and female students' responses to the CBAS questionnaire across countries 54

Table 18. Relative effort across genders and tests 59

Table 19. Female and male students' preference for the CBAS or paper-and-pencil test across countries 60

Table 20. Correlations between CBAS enjoyment, motivation and achievement 63

Table 21. Correlations between paper-and-pencil enjoyment, paper-and-pencil motivation and paper-and-pencil achievement 64

Table 22. Average reported effort per country for the CBAS and the paper-and-pencil tests 69

Table 23. Average reported effort if the CBAS and the paper-and-pencil tests had counted for school marks 70

Table 24. Correlations between reported effort and achievement 72

Table 25. Correlations between item order, item difficulty and proportion of missing responses 76

Table 26. Items on which females had more missing responses than males 77

Table 27. Items on which males had more missing responses than females 78

Table 28. Item-based word count for CBAS and PISA items 80

Chapter 1

1

Introduction

Computer-based assessments are becoming more and more prevalent in the educational testing domain and the computer-based assessment of science in PISA was envisioned to test a number of important aspects of this methodology in comparison to paper-and-pencil methods. Only three of the PISA countries, Denmark, Iceland and Korea, finished this project and those results are reported here.

Key findings from the PISA 2006 Computer-based Assessment of Science

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

Males outperform females 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.

WHAT IS PISA?

PISA (Programme for International Student Assessment) is a triennial survey of the knowledge and skills of 15-year-olds commissioned by the Organisation for Economic Co-operation and Development (OECD). The PISA test of reading, mathematics and science is the most comprehensive and rigorous international programme to assess student performance and to collect data on the student, family and institutional factors that can help to explain differences in performance. Decisions about the scope and nature of the assessments and the background information to be collected are made by leading experts in participating countries and are steered jointly by governments on the basis of shared, policy-driven interests. Stringent quality assurance mechanisms are applied in translation, sampling and reliability, and can significantly improve understanding of the outcomes in the world's economically most developed countries, as well as in a growing number of countries at earlier stages of development.

More than 400 000 students from 57 countries took part in the PISA 2006 assessment. The focus in this assessment cycle was on science literacy and the assessment included an optional computer-based component assessing scientific competencies. This report focuses on the results from that computer-based component (Computer-based Assessment of Scientific Literacy - CBAS) which three countries administered (Denmark, Iceland and Korea) and compares their computer-based achievement results to results obtained on the PISA 2006 paper-and-pencil assessment of science. More information about the PISA approach to reading, mathematical and scientific literacy and the cross-country findings from the paper-and-pencil PISA 2006 assessment can be found in the PISA publications at *http://www.pisa.oecd.org*. The tables and charts in this report are based on the PISA International Computer-based Assessment Database, available at *https://mypisa.acer.edu.au/*.

WHAT IS A COMPUTER-BASED ASSESSMENT OF SCIENCE?

Delivery of student assessments via computer is becoming more and more prevalent in the educational testing domain as changes are made in assessment methodologies that reflect practical changes in pedagogical methods. In Iceland, projects are underway to change the current paper-and-pencil national testing platform to a computer-based platform that will be delivered over the Internet and in Denmark, the full range of national student assessments have been migrated recently to a computer-based platform. Introducing computer-based testing is viewed as a necessary and positive change in educational assessment – not only in that it reflects changes in classroom and self-teaching methods that are more and more computer-mediated – but also in that it can provide a number of assessment and administration advantages.

Computer-based assessment:

- requires fewer language skills;
- can be easier and less costly to administer for schools in that there are no hard copy tests to be printed, distributed and returned to the testing institute;

- facilitates complex test design, for example, students in the same test setting can easily be assigned to different versions of the assessment;
- can present more information succinctly and in a shorter space of time;
- can collect more information, such as time taken per item, key strokes made and a student's movement through the questions;
- can score electronically directly from student responses, freeing up teacher time usually used for marking tests;
- is particularly useful in the assessment of science for simulating scientific phenomena that cannot easily be observed in real time such as seeing things in slow-motion or speeded-up, for modelling scientific phenomena that are invisible to the naked eye (*e.g.* the movement of molecules in a gas), for presenting students with the opportunity to perform repeat trials in limited testing time, or for working safely in lab-like simulations that would otherwise be hazardous or messy in a testing situation; and
- can lead to assessments that have a dual goal of assessing while also assisting learning (for example, by providing 'clues' to students whose first attempt at the question is incorrect).

This last advantage is particularly important in Iceland where there is high support for a move towards more individualised testing. The Educational Testing Institute recently made a proposal to change the national testing system in Iceland over to a computerised and adaptive testing system and this has been received with enthusiasm by teachers, politicians, schools and parents. This move can be seen as a natural progression in Iceland's long history of national testing, given the extremely high use of the Internet and computers in Iceland. Both the PISA 2003 and the PISA 2006 results show that Iceland has the highest rate of 15-year-olds with Internet access at home (97.7% in PISA 2006).[1] The PISA 2006 results also show that Iceland has the second highest rate (after Denmark) of computer access in the home (98.1%) (OECD, 2004, 2007a).

Given the rapidity with which these methods of student assessment are being implemented in a number of settings, current research needs to concentrate on understanding the impact on performance for students taking on-line or computer-based assessments compared to paper-and-pencil assessments and to design computer-based assessments so that students have the best opportunity to show their true level of proficiency.

Currently, many questions still remain unanswered in terms of the impact of changing testing systems over from paper-and-pencil to computer-based and should be considered before major changes are made to assessment systems. For example, 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? Does a move to on-screen tests affect male and female test-takers in the same way? Can the same types of competencies be assessed via computer as by paper-and-pencil methods? Does the move to computer-based testing affect students in the same way across various subject areas such as reading, mathematics and science?

Previous research has in part indicated the psychometric equivalence of computer-based and paper-and-pencil tests (Mason, 2001; Singleton, 2001; Wilhelm & Schroeders, 2008; Zandvliet & Farragher, 1997). On the other hand, 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. Gallagher, Bridgeman and Cahalan (2000) suggested that females may experience an increase in stereotype threat

under the computer-based condition which negatively affects their performance. They reason that "because some differences in impact due to testing format (paper-based or computer-based) occur consistently across several tests, we can assume that changes in impact are due to changes in format and not to anomalies associated with changes to specific test" (p.15). The OECD report on Information and Communication Technology in PISA assessments goes further to state that "the more advanced the task, the wider the (gender) gap" (p.50, OECD, 2005b).

Suggestions have been made that males may outperform females on computer-based tests because males generally show a higher degree of ICT familiarity than females (OECD, 2005b). Conversely, other research has shown that ICT familiarity differences do not affect performance (Taylor, Jamieson, Eignor & Kirsch, 1998). In terms of test enjoyment for computer-based tests, studies have consistently shown that students prefer computer-based tests to paper-and-pencil tests (Singleton, 2001; Zandvliet & Farragher, 1997), though the impact of greater enjoyment on achievement remains to be confirmed. Additionally, analysis of the impacts of computer-based test type and configuration requires further investigation. A report by the Educational Testing Service (Bridgeman, Lennon & Jackenthal, 2001) indicated that test-takers using high resolution screens scored higher on a verbal test than test-takers using low resolution screens.

However, a note of caution for interpreting these results is necessary here. When interpreting results from previous research, it is important to keep in mind the extremely rapid and largely unmonitored changes occurring in the nature of students' computer use. Types of ICT activities, access to the Internet and time spent on the Internet are rapidly changing in almost every country across the world. In the past year prior to publication of this report social networking sites popular with high school students such as Facebook and LinkedIn seemingly materialised overnight and now have over 70 million users, and an additional 150 000 users per day (Hoffmann, 2008). Other sites such as MySpace have grown exponentially in popularity since their introduction approximately five years ago, as, due to improvements in technology, the Internet has become increasingly accessible around the world and people, including 15-year-olds, spend more and more time on it.

Students' computer use takes place, for the most part, at their homes in their spare time and the necessary skills are self-taught or learnt from peers. As a result, it is very difficult for researchers to monitor students' out of school use due to the *i*) rapidly increasing frequency of use, *ii*) the changing nature as new trends such as Facebook are introduced and almost simultaneously adopted by millions of users around the world, and *iii*) students teach themselves or their peers the skills necessary to use these technologies. As a result, many of their activities may 'pass under the radar' of parents, teachers and researchers alike.

Data from PISA 2003 (OECD, 2004) can give an indication of just how much student learning of computer skills remains within their peer group.

Table 1 shows the responses to the question that asked students where or from whom they learnt the most about how to use computers and the Internet. The two most common responses are listed here.

The trends data from the last three cycles of PISA support the notion that it may be of little use to take results investigating students' computer use from the past (even from the last five years) and generalise it to today's 15-year-olds. Students' frequent use of computers and the Internet is increasing rapidly and may also be changing in nature as new activities are introduced such as blogs, Internet chat and social networking. These results are displayed in Figures 1, 2 and 3 where the percentages of students reporting that they have a computer at home, have a computer with Internet access, and use the computer almost every day have increased at a dramatic rate over the past nine years in the three countries of interest in this report as well

Table 1.
Sources of learning about computers and the Internet (PISA 2003)

	How I learnt to use the Internet		How I learnt to use computers	
	By myself	From friends	By myself	From friends
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%

*OECD Average - mean data for all OECD countries with each country contributing equally to the average

as in the OECD overall. (Note that Iceland and Korea did not participate in the Information Communication and Technologies questionnaire in PISA 2000 so their data cannot be displayed in Figure 3 or Figure 4.)

Figure 4 shows that this pattern of rapid and substantial increases in the frequency of computer use is not shared across all types of computer use as, in contrast to use of computers in the home, use of computers at school appears relatively stable over time, in the three countries of interest for CBAS and across the OECD as a whole.

Figure 1.
Proportion of students possessing a computer in the home over PISA cycles

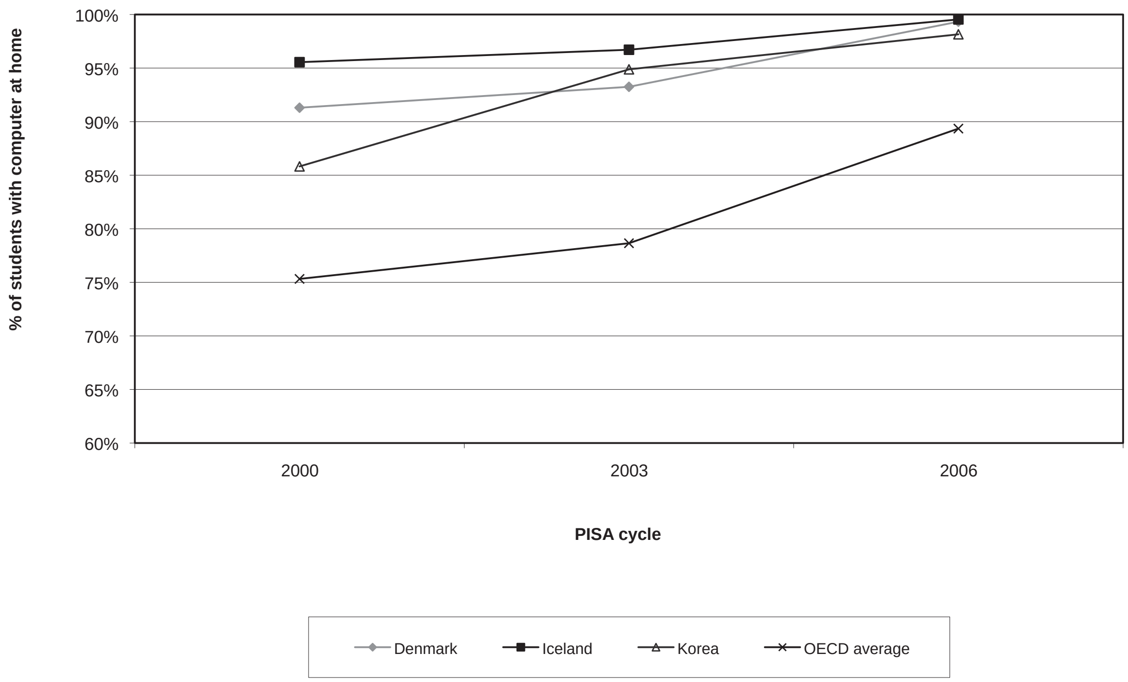


Figure 2.
Average percentage of students with Internet connection at home across PISA cycles

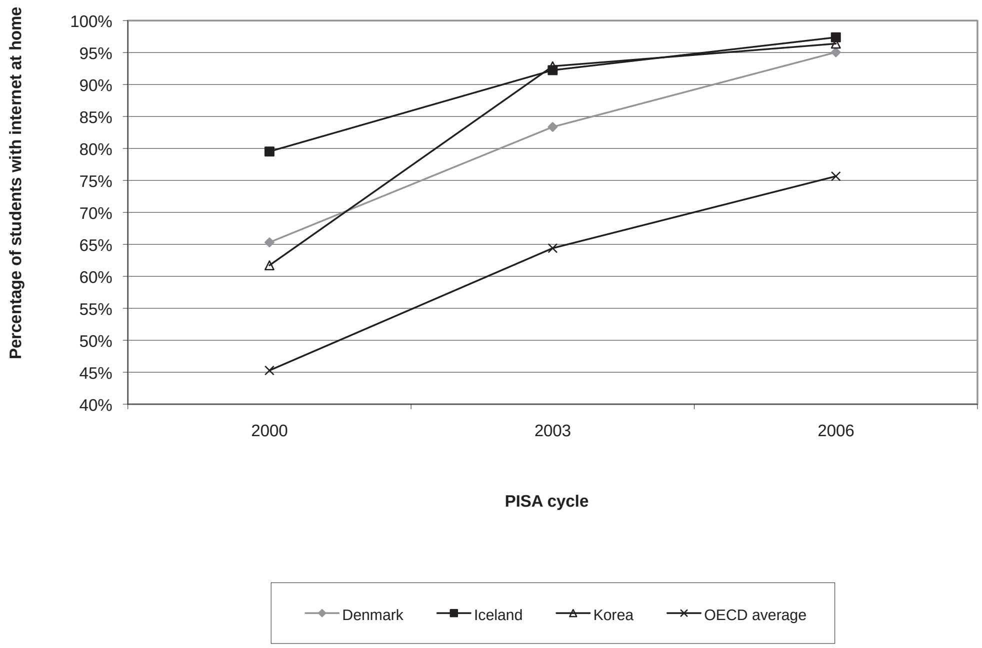


Figure 3.
Average percentage of students using a computer at home almost every day

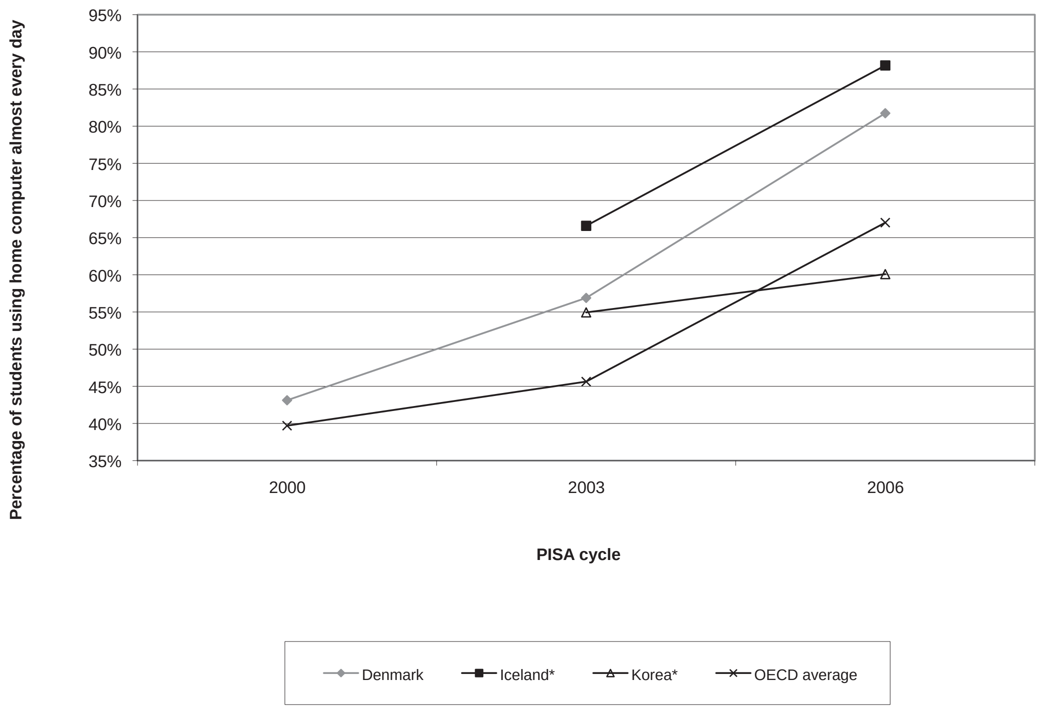
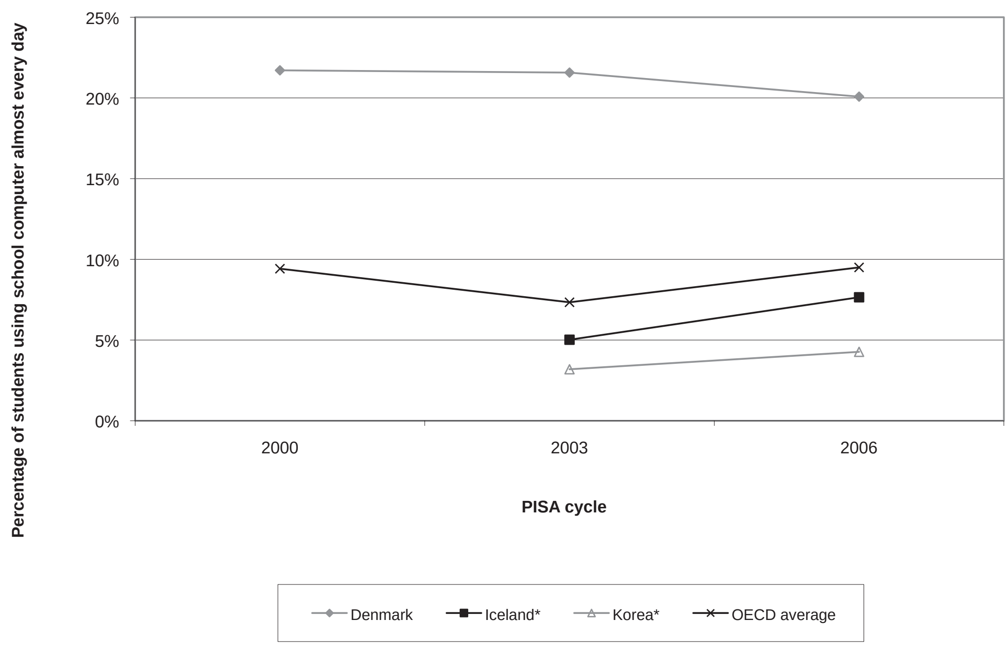


Figure 4.

Average percentage of students using a computer at school almost every day



OBJECTIVES OF THE PISA 2006 COMPUTER-BASED ASSESSMENT OF SCIENCE (CBAS)

The initial aim of the extension of the PISA 2006 science assessment to include a computer-delivered element was to administer questions that would be difficult to deliver in a paper-and-pencil test – the relevant questions included video footage, simulations and animations. This also reduced the amount of reading required so that students' science capacity was assessed more directly and the results could more accurately reflect their true ability. The development of a computer-based component helped with the drafting of PISA science questions and the creation of several procedures including those of automatic coding.

Notes

1. Korea and Denmark, the other two countries examined in this report are ranked 5th and 8th at 96.5% and 95.5% respectively of 15-year-old students having Internet access at home.

Chapter 2

2

Administration of the CBAS test

The computer-based assessment of science was field tested in thirteen PISA countries in 2005 and the main study was carried out in three of them in 2006, Denmark, Iceland and Korea. Rigorous sampling was used in all countries and all students also took the regular PISA test. The computer-based test had appropriate test characteristics and sampled a number of different aspects of science knowledge and skills.

FIELD TRIAL

CBAS was administered in the PISA 2006 Field Trial in 13 PISA-participating countries. Preliminary analyses were performed on the data from 12 of these countries to determine the final item selection and final method of test administration for the Main Study which was undertaken later in 2006. Only three countries (Denmark, Iceland and Korea) continued through to the Main Study phase of CBAS and the data from these countries are the focus of this report.

SCHOOL AND STUDENT SAMPLING

A subsample of 100 schools was selected to participate in CBAS from the main PISA 2006 school sample test in Iceland. From these schools, clusters of 5 to 45 PISA-eligible students were sampled from the PISA student sample. All schools and students selected for CBAS had already participated in the paper-and-pencil PISA 2006 assessment.

It is important to note that the sample considered in the present analyses includes all students that participated in the CBAS test session as well as all PISA-participating students from the schools that had at least one student participating in CBAS. For Iceland, the original CBAS sample was drawn with 1 104 students out of which 784 students participated (71% response rate). However these analyses include data for an additional 2 782 students who participated in the paper-and-pencil assessment of science, attended a CBAS-participating school but did not respond to the CBAS test. To give achievement scores on the CBAS test for these students, plausible values on the CBAS scale were statistically imputed based on the students' PISA paper-and-pencil achievement and background information. A total of 3 566 students are therefore included in the CBAS analyses for Iceland, which is very close to the total number of students participating in the paper-and-pencil PISA 2006 (3 789). As a result, it is fairly certain that the Icelandic sample for CBAS is representative of the population of 15-year-old students in the country.

In Denmark and Korea, approximately half of the PISA sampled schools were sampled for CBAS and, again, all students participating in CBAS or PISA from these schools were included in the final data file. The final numbers of included students were: 837 CBAS-responding students out of 1 254 students included in the CBAS analyses for Denmark and 1 475 CBAS-responding students out of 2 650 students included in the analyses for Korea.

Sample weighting

To account for any biases in selection of schools and students, the PISA data are weighted using a balanced repeated replication method. This accounts for, for example, any over- or under- representation of geographical areas within countries. More information about the weighting techniques in PISA can be found in the Data Analysis Manual (OECD, 2005a) or in the PISA 2006 Technical Report (OECD, 2009).

Gender distribution in the weighted sample

In Denmark and Iceland the CBAS sample was approximately equally constituted of males and females. Table 2 shows however, that there was a greater number of males than females in Korea. The greater number of males for Korea is of concern given that Field Trial results indicate a strong gender difference in achievement in favour of males. This should be kept in mind when interpreting cross-country achievement comparisons.

Table 2.
Proportion of females and males in the sample analysed in this report

	CBAS sample	
Country	Females	Male
Denmark	52%	48%
Iceland	50%	50%
Korea	44%	56%

ADMINISTRATION OF TEST SESSIONS

CBAS sessions took place either on the same day as the PISA paper-and-pencil assessment of students' reading, mathematics and science performance, or very shortly thereafter. Test administration was standardised so that all students performed the test on the same type of laptop, using the same software and in a similar testing environment.

Up to five students participated in each test session under the guidance of one Test Administrator. The computer-based science items were presented to students on laptop computers through CBAS software specially designed for this purpose. This was a fixed-form test where the same 45 items were presented to all students in one of either two orders. The order of items was split from the middle point of the second form so as to reduce fatigue effects on the items occurring later in the test.

The software allowed students to move between items as they wished and to return to questions (changing their answers if necessary) up until 55 minutes had elapsed since the beginning of the test, at which point the Test Administrator stopped the session. This allowed for just over one minute per question. If a student finished early the items remained on the screen until the completion of the 55 minute test session.

Following the cognitive items, four questionnaire items were presented and students had five minutes to respond to these. In total therefore, test sessions were one hour long.

Technical specifications

All laptops used for student testing were required to comply with a number of minimum specifications as given in Table 3 below.

Table 3.
Minimum hardware specifications for student laptops used for CBAS

Component	Minimum required
Central processing unit	1.6 GHz Intel Pentium® M Processor
Memory	512 MB of RAM
Hard disk	40 GB
Display	14.1″ XGA (this exact screen size must be used)
Network	Must be equipped with wireless capability (Centrino technology) – IEEE 802.11 b/g
Optical drive	Computers must be equipped with a DVD-ROM drive for installing the software required for test administration. Also, the test administrator's computer must be equipped with a CD-RW drive for writing test session result data to CDs.
Operating System	Windows® XP Professional
Pointing device	Symmetrical external, optical mouse
Listening device	External stereo headphones

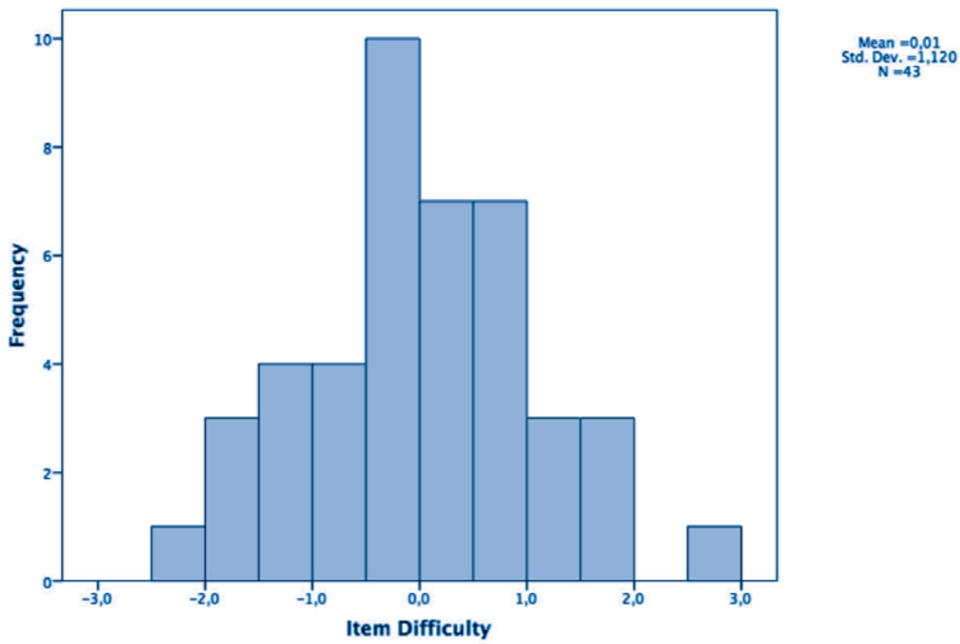
Cognitive item information

In total, 45 items with multimedia extensions (animations using flash software, video footage or photos) were presented to students. The final analyses are performed on 42 items as two items (Spinifex Q2 and Assembly Line) were dropped prior to the analyses and two items were combined into one (Bicycle Q1 and Q2) as they were considered to be assessing the same knowledge. Two additional items were set to 'not administered' for Icelandic students (Tablet Fizz Q2 which shows video footage of a vitamin tablet dissolving in water which was judged as an unfamiliar concept for Icelandic students and Fish Farm Q2 because one of the specific terms in the item was not translatable into Icelandic). 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 order and missing responses

The two items that were removed from the database prior to the analyses for all countries may have been affected by their late presentation in both forms: one was the final question on Form 1 and the 23rd item on Form 2 (and had the highest combined order rank score) and the other was the second to last question on Form 2 (40th out of 45 in order rank). This indicates that fatigue may have played a substantial role in the way students responded to these items which eventually saw them deleted. Missing responses also increased towards the end of the test, however, overall the proportion of missing responses was very low for all countries (less than 1%). Further examination of gender differences in missing responses will be presented in a later section of this report.

Figure 5.
Distribution of item difficulty for final CBAS items



Item difficulty

Item difficulties were calculated and Figure 5 above shows that the items were approximately evenly distributed across the item difficulty scale from -3 to 3 with the mean item difficulty at zero, indicating good coverage of all potential competency levels. More information about the calculation of PISA item difficulties can be found in the 2006 PISA Technical Report (OECD, forthcoming). Percentage correct per CBAS item was also calculated and ranged from 13% to 94% with an average percentage correct per item of 60%.[1] Percentage correct per item was strongly associated with item difficulty from the model at 0.90 indicating that percentage correct per item is also an adequate measure of performance for specific analysis purposes.

Item difficulty across genders

After the Field Trial, to determine the final selection of Main Study items, item difficulties were calculated per gender in a gender differential item functioning (DIF) analysis. Two of the items identified as easier for females were maintained in the Main Study (Ping Pong Ball Q1 and Washing Glasses Q3) and three items that were identified as easier for males were maintained in the Main Study (Bicycle Q2 and Q3 and Designed for Flight). Performance on these items will be examined in closer detail in the results section of this report.

Science framework classification

In the same way that paper-and-pencil science items were classified into domains of science, CBAS items were categorised according to one of the three scientific competencies of PISA (Explaining Phenomena Scientifically, Identifying Scientific Issues or Using Scientific Evidence) and according to what sort of knowledge was being assessed. They were either considered to be assessing:

1. Knowledge of Science (*i.e.* knowledge about the natural world, about the topics of science) and one of the sub domains of knowledge: Physical systems, Living systems, Earth and Space systems or Science and Technology, or,

2. Knowledge about Science (*i.e.* knowledge about science itself and the methodologies of science) and one of the sub-categories: Scientific Enquiry or Scientific Explanations.

The final CBAS items selected for the Main Study were relatively evenly distributed across the types of scientific knowledge and sub domains of science and were similar in distribution to the paper-and-pencil science items. Slight differences were found in the proportions of items assessing the *explaining phenomena scientifically, identifying scientific issues* and *using scientific evidence* between tests in that the CBAS test had more items assessing (presumably because the computer medium facilitated items involving scientific trials and repetition more so than did the paper-and-pencil medium). In terms of domains, the CBAS test had slightly more technology items and items involving the physical sciences whereas the paper-and-pencil test had more items assessing living systems. However, as the distributions across categories in Table 4 show, these differences are relatively slight. Overall, it seems that the paper-and-pencil test and the computer-based test were assessing similar areas of scientific competencies and knowledge.

Table 4.
Distribution of paper-and-pencil science and CBAS items across domains of science

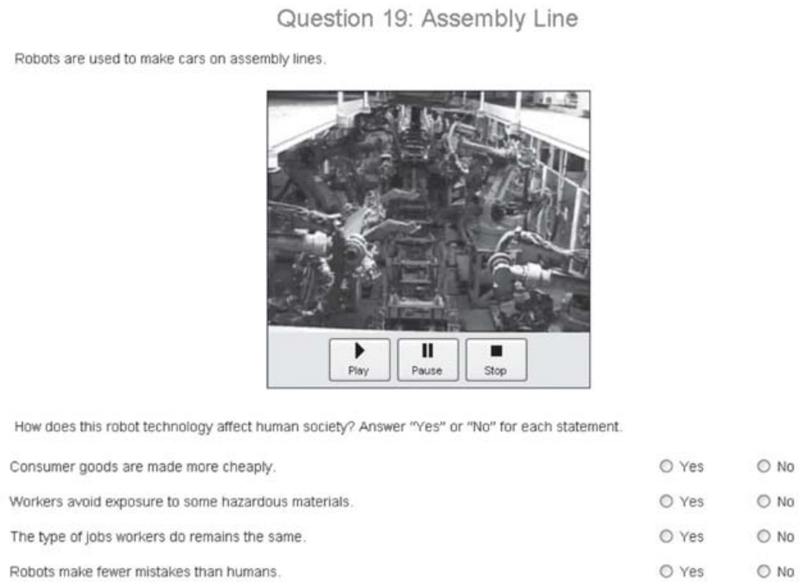
			Paper-and-pencil Science	CBAS
Scientific Competency		Explaining Phenomena Scientifically	49%	40%
		Identifying Scientific Issues	22%	24%
		Using Scientific Evidence	29%	36%
Scientific Domain	Knowledge of Science	Earth and Space Systems	11%	12%
		Living Systems	23%	17%
		Physical Systems	16%	21%
		Technology Systems	7%	10%
	Knowledge about Science	Scientific Enquiry	23%	19%
		Scientific Explanation	19%	21%

Interactivity

As the computer-based items differ markedly from the paper-and-pencil items in terms of how much the student can interact with the item (for example, the possibility of moving levers to adjust levels in experimental trials or dragging and dropping the answer into the correct location in the diagram), it may be useful to consider the effects of the interactivity of the items on performance, in particular across genders.

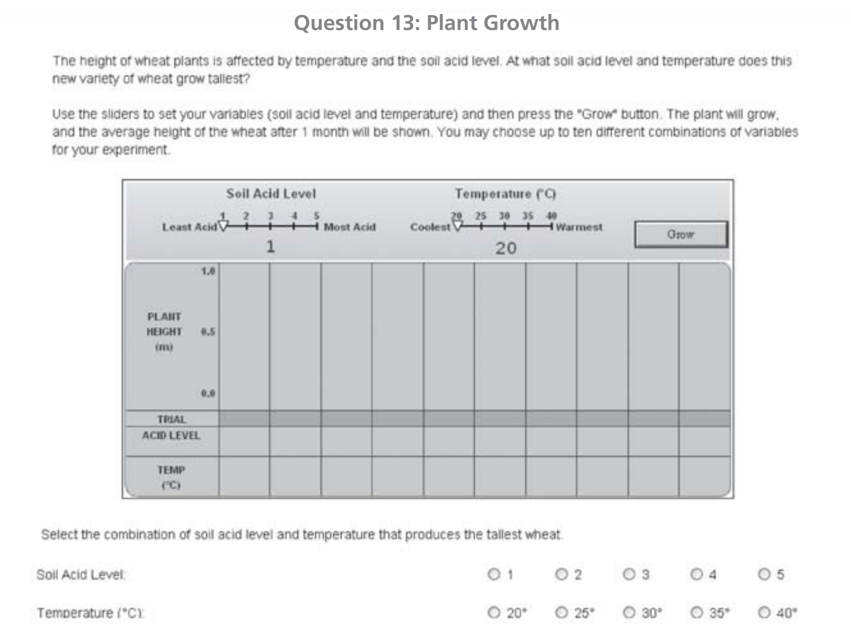
A panel of three independent judges rated all CBAS items into three groups according to the level of interactivity (low, medium and high) based on the types of activities the student had to perform with the item and based on how much the student needed to engage with the audiovisual material to answer the question. An example of a low interactivity item is the "Assembly Line" item in Figure 6 that shows a short video of an automated car assembly line and asks a question related to the role of robots in society.

Figure 6.
Sample unit: Assembly Line



Here, the video footage serves as contextual information to the item but does not provide the answer. In fact, this question could be answered correctly without the student watching the video footage and is therefore considered to be of low interactivity. In contrast, the following item in Figure 7, "Plant Growth", where the student is required to move buttons up and down a scale, performing experimental trials on optimal temperature and soil acidity levels for growing wheat, was considered as highly interactive.

Figure 7.
Sample unit of highly interactive item: Plant Growth



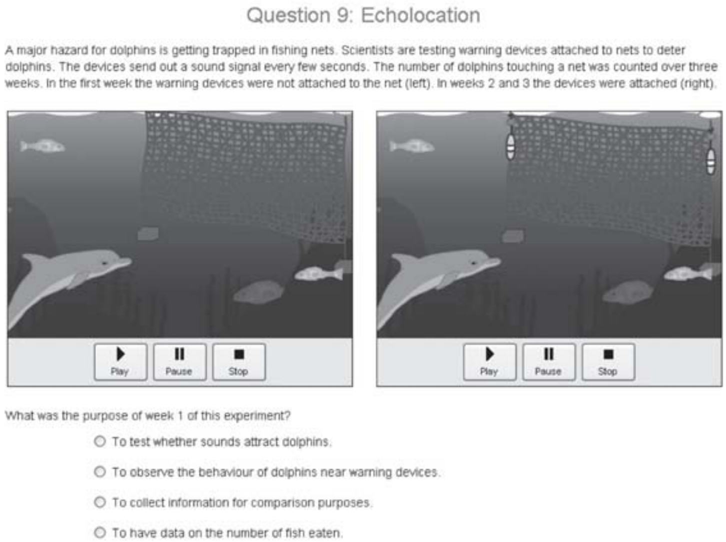
Overall, 14 items were classified as high interactivity, 13 as medium and 16 as involving low interactivity. Achievement on these items will be compared in the results section.

Reading load

Word counts for each CBAS item were recorded according to the number of words embedded in the image of each stimulus in both the question stem and multiple choice response options. Based on these figures, the CBAS items were divided into three groups according to reading load: low, medium and high. Eleven items were considered to be of a high reading load, for example as shown in Figure 8.

Figure 8.

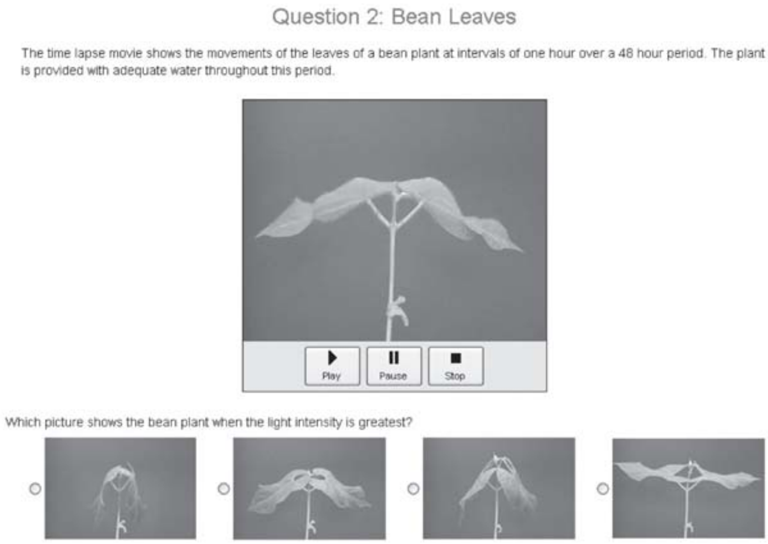
Sample item showing high reading load item: Echolocation



Fourteen items were classified as medium reading load and eighteen items (including the item in Figure 9 below) were classified as low reading load.

Figure 9.

Sample item showing low reading load item: Bean Leaves



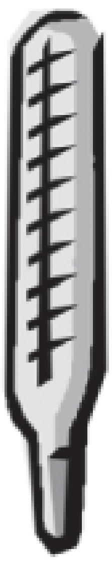
CBAS questionnaire items

After the cognitive items the students were asked to respond to several short questions to investigate the effects of enjoyment, motivation and effort on performance. Students were asked to use a four-point Likert scale to rate how much they enjoyed the computer-based and paper-and-pencil tests, and whether they would do a similar test where the answers were provided "just for fun" (assessing motivation). The PISA Effort Thermometer was also used where students were asked to imagine an actual situation that was highly important to them personally, so that they would try their very best and put as much effort as they could to do well. They were told that in this situation they would mark the highest value on the effort thermometer (10) and then they were asked to report: how much effort they put into doing the CBAS test compared to the situation they had just imagined; and how much effort they would have invested if their marks from CBAS had counted in their school marks. This questionnaire item (identical to the item used in the PISA paper-and-pencil test) is displayed below in Figure 10.

Figure 10.
PISA Effort Thermometer

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?
☐ 10	☐ 10	☐ 10
☐ 9	☐ 9	☐ 9
☐ 8	☐ 8	☐ 8
☐ 7	☐ 7	☐ 7
☐ 6	☐ 6	☐ 6
☐ 5	☐ 5	☐ 5
☐ 4	☐ 4	☐ 4
☐ 3	☐ 3	☐ 3
☐ 2	☐ 2	☐ 2
☐ 1	☐ 1	☐ 1

In addition, students were asked which test they put more effort into between the CBAS test and the PISA paper test (assessing relative effort) and they were finally asked what type of test they would prefer between a two hour paper-and-pencil test, one hour of each type of test and two hours of computer-based testing.

Notes

1. Percentage correct for all PISA Science items was calculated across the three CBAS participating countries, ranging from 9% to 91% with an average percentage correct across items at 46% which is lower than that for CBAS. Overall, percentage correct per paper-and-pencil science item was correlated with the item difficulties at 0.992 indicating that percentage correct is also an adequate indication of performance.

Chapter 3

3

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. In the computer based test, male performance increased in Iceland and Korea while female performance decreased. Males outperformed females on the computer-based test in all three countries. Females outperformed males on the paper-and-pencil test of science in Iceland whereas there was a gender difference in favour of males in the paper-and-pencil results for Denmark. The association between reading literacy and achievement on the science test was weaker for the computer-based items than for the paper-and-pencil items.

ACHIEVEMENT ACROSS TEST MODALITIES

Scaling information

Initially, CBAS scores for the three countries were scaled on the traditional PISA scale with a mean of 500 and a pooled standard deviation (SD) of 100. Paper-and-pencil science, reading and mathematics scores for the three CBAS countries were also re-scaled from the same model as the CBAS plausible values so as to allow calculation of correlations between CBAS and the paper domains. (More information about this technique is given in Annex A.)

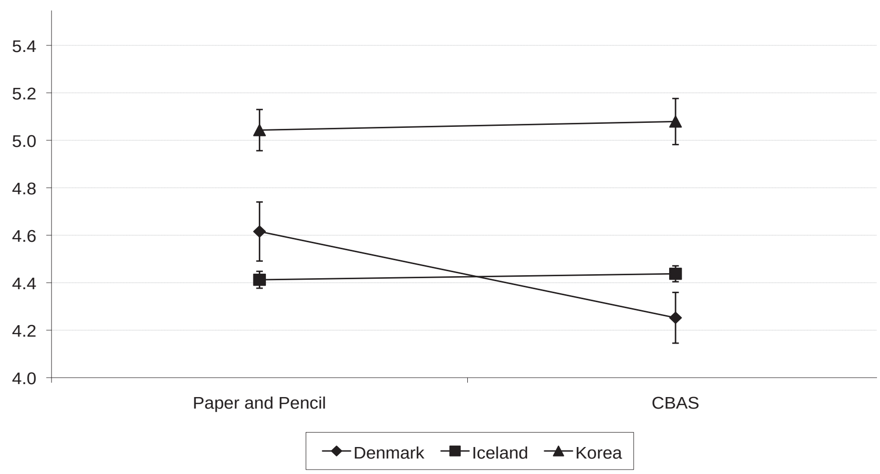
Because these new scores were re-scaled for only three countries, they are not directly comparable with the OECD-reported PISA 2006 test scores where 500 and 100 are the mean and SD of the 30 OECD countries. Therefore, to avoid confusion between the scales, in this report all achievement scores have been re-standardised on a new scale with a mean of 5 and a SD of 2, meaning that over 99% of students have scores between 0 and 10 on the scale. This removes the possibility of direct comparisons between the scores reported here and the scores reported in the OECD PISA 2006 report which would not be valid (because the plausible values are drawn from different models).

Performance across countries and test modalities

As the CBAS items are normally distributed around a mean of zero, which is similar to the item difficulty of the set of paper-and-pencil science items, it is assumed for the following analyses that the CBAS and the paper-and-pencil science tests were of equal difficulty (the mean for both tests is also equally set at 5 and equally distributed with a SD of 2). This means that differences between the two performance scores can be interpreted as impacts of the features of the test modality; that is, the type of computerised items, presentation on the screen, interactivity, lower reading load, etc.

As the following Figure 11 shows, performance on the paper-and-pencil science and CBAS tests was higher in Korea than in both Denmark and Iceland. These results are similar to the PISA 2006 reported results for scientific literacy where Korean students outperformed Danish and Icelandic students. The difference here may be slightly artificially inflated however, due to the slightly higher proportion of males in the CBAS sample for Korea but the difference is nevertheless substantial. There were no differences between Denmark's and Iceland's performance on CBAS, nor on the paper-and-pencil test. Whereas Korean and Icelandic students' performance remains relatively stable across test modalities, the data indicate a trend for an overall drop in performance for Danish students on the CBAS test, although this is not statistically significant.

Figure 11.
Achievement in paper-and-pencil test of science compared to CBAS[1]



Strength of relationship between paper-and-pencil scores and CBAS scores

As Table 5 and Figure 12, Figure 13 and Figure 14 show, science achievement on the computer-based test and on the paper-and-pencil test was highly correlated overall. The correlations in Iceland were somewhat weaker than in Denmark and Korea, perhaps because of the size of the gender difference reversal for CBAS. No significant differences in the strength of the correlations are observed across genders.

As previously noted in OECD PISA reports (OECD, 2007a), reading and science achievement is highly correlated, although specifically of interest for this report is the relatively weaker correlation between CBAS and reading achievement (0.75) compared to the correlation between paper-and-pencil science and reading achievement (0.83). This indicates that achievement on CBAS was less dependent on reading ability than achievement on the paper-and-pencil test. As a result, one can conclude that one of the original aims of CBAS has been met in that the PISA computer-based test is less confounded by reading ability than achievement on the paper-and-pencil test.

Table 5.
Correlations for male and female students' performance in CBAS and paper-and-pencil tests across countries

	CBAS and science paper-and-pencil scores			CBAS and reading paper-and-pencil scores			Reading paper-and-pencil and science paper-and-pencil scores		
Country	Female	Male	Total	Female	Male	Total	Female	Male	Total
Denmark	0.88	0.90	0.89	0.82	0.78	0.77	0.82	0.79	0.79
Iceland	0.78	0.80	0.78	0.72	0.73	0.68	0.85	0.88	0.86
Korea	0.88	0.90	0.88	0.77	0.78	0.74	0.85	0.85	0.84
Total	0.88	0.90	0.88	0.78	0.78	0.75	0.84	0.84	0.83

Figure 12.
Average CBAS score and average paper-and-pencil science scores in Denmark (N=1254)

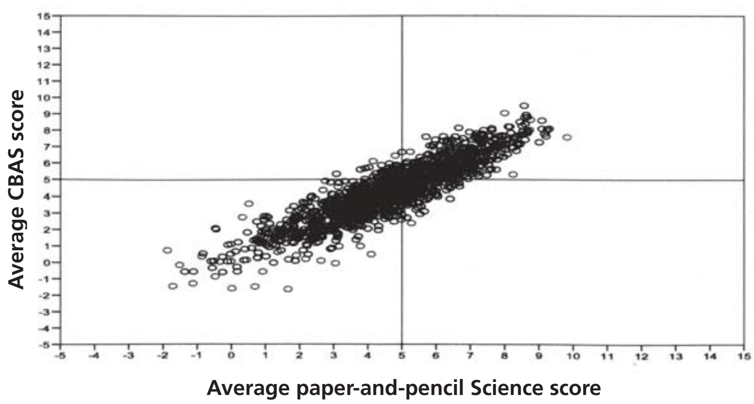


Figure 13.
Average CBAS score and average paper-and-pencil science scores in Iceland (N=3566)

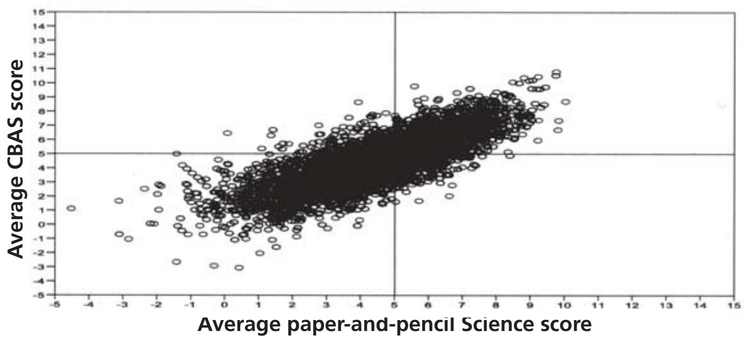
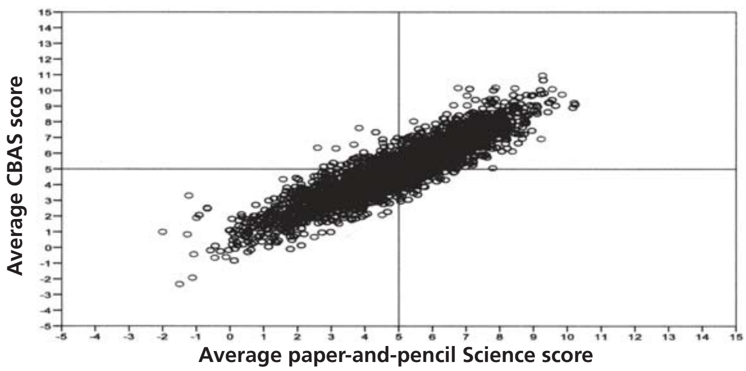


Figure 14.
Average CBAS score and average paper-and-pencil science scores in Korea (N=2650)



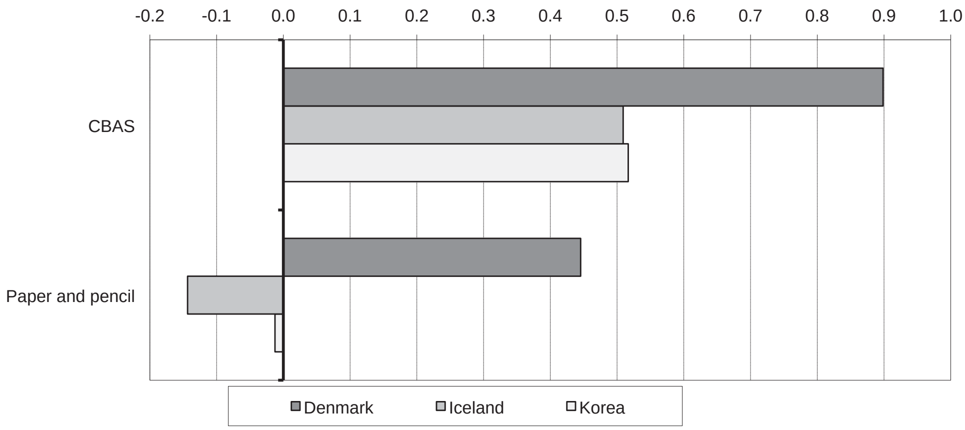
Gender differences in student performance across test modalities

As Figure 15 below shows, in Denmark, males performed significantly better than females on the paper-and-pencil test of science by almost ¼ of a standard deviation. In Iceland, females slightly outperformed the males on the paper-and-pencil test of science and in Korea there were no significant gender differences. The gender differences were large and clearly directional when science achievement was tested via computer however, with males performing better than females on the CBAS test in all countries.

Males outperformed females on CBAS by approximately ¼ of a standard deviation in Iceland and Korea to almost half a standard deviation in Denmark. Denmark has the largest gender difference in favour of males regardless of test modality, but it should be noted that the increase in size of the gender difference as students moved from one test to the other is similar across all three countries. In other words, in Denmark the gender advantage for males increased by ¼ of a standard deviation from ¼ to ½. In Korea it also increased by ¼ of a standard deviation from 0 to ¼ and in Iceland the increase was slightly larger as the advantage was reversed from 1/10 of a standard deviation to ¼ of a standard deviation.

Figure 15.

Male achievement advantage across tests and countries (positive values show males outperforming females)



When the mean achievement scores for females and males are compared in the paper-and-pencil test to the CBAS test across countries in Table 6 shows that in Denmark males' and females' CBAS performance drops (with female performance dropping more than males increasing the gender difference). In Iceland and Korea, males' performances increased, overtaking the females whose respective performance decreased and creating the gender difference seen earlier. Statistically significant differences are tested with the means and standard errors of the mean calculated through the replicates procedure involving the eighty PISA replicate weights on plausible values. When a gender difference is statistically significant at the p<0.05 level of significance, the male and female means have been printed in bold in the table below.

Table 6.

Achievement in paper-and-pencil test of science compared to CBAS

	Paper-and-pencil[2]			CBAS		
Country	Females	Males	Total	Females	Males	Total
Denmark	**4.39** (0.16)	**4.85** (0.14)	4.62 (0.12)	**3.81** (0.15)	**4.71** (0.12)	4.25 (0.11)
Iceland	**4.49** (0.05)	**4.34** (0.05)	4.41 (0.04)	**4.18** (0.04)	**4.69** (0.05)	4.44 (0.03)
Korea	5.06 (0.13)	5.03 (0.12)	5.04 (0.09)	**4.79** (0.14)	**5.31** (0.13)	5.08 (0.10)

(Significant differences are shown in bold)

The correlations in Table 7 below further show that female CBAS performance is slightly less strongly associated with their performance on the paper-and-pencil test of science than for males, indicating that the impact of changing the test method is not the same for females as it is for males.

Table 7.

Correlations between paper-and-pencil science scores and CBAS scores across genders and countries

Correlations	Females	Males
Denmark	0.89	0.91
Iceland	0.78	0.80
Korea	0.88	0.90

Notes

1. Note that the error bars represent the 95% confidence intervals for the mean achievement score. Strictly, these cannot be used to test for significant differences between groups however as this requires calculation of the standard error of the mean taking into account the five plausible values applying the eighty replicate weights. For significance testing the PISA replicate macros in SPSS were used. See the Data Analysis Manual for more information about this procecure (OECD, 2005a).

2. Note that the gender difference in Iceland on the paper-and-pencil test of science is in favour of the females. This was the trend reported in the PISA 2006 OECD report, however with the full sample this difference was not statistically significant. Analysing only the students included in the CBAS sample however, this difference becomes significant at the $p<0.05$ level.

Chapter 4

4

Use of Information Communication Technologies (ICT) across genders and impact on achievement

The vast majority of students have computers and Internet connections at home. Not having a computer at home is associated with poorer performance on the CBAS test. Overall, males score higher than females on the frequency of use of scales. Perhaps as a result of this, they also score higher on the confidence in using ICT scales, particularly in Iceland.

Females tend to use the Internet more for social networking activities such as chatting and email and their confidence is consequently higher for these activities. Males tend to browse the Internet, play games and download software a lot more than females and they perform advanced computer activities more frequently. Overall, males have much more confidence than females in most ICT activities, which is also found in ICT PISA 2003 results.

The following section examines the relationship between familiarity with and confidence in using ICT and achievement. Previous studies have indicated that use of computers in the home (and greater ICT confidence) is strongly correlated with higher academic achievement (Harrison, Comber, Fisher, Haw, Lewin, Lunzer, McFarlane, Mavers, Scrimshaw, 2003; Ravitz, Mergendoller & Rush, 2002). Further research specifies that only home use of computers for educational purposes was associated with higher performance (in mathematics), whereas out-of-school use of ICT was negatively associated with performance (Valentine, Marsh, Pattie and BMRD, 2005). The OECD PISA 2003 ICT report revealed that higher scores on the Internet use and Programme scales were actually associated with a drop in mathematics and reading performance, stating that "one cannot readily assume that computer usage is bound to be beneficial for students in all cases" (OECD, 2005b, p.65).

HOME ICT ACCESS AND ACHIEVEMENT IN CBAS

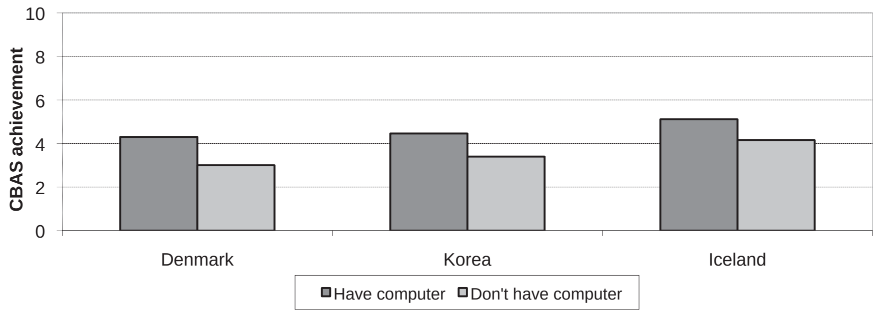
As Table 8 shows, the vast majority of students in the CBAS-participating countries had access to a computer in the home with an Internet connection. Access to educational software was less common, but the majority of students in each country still reported that they had access to educational software in the home.

Table 8.
Proportion of students in each country with a computer, educational software and Internet at home

	Have computer at home	Have educational software at home	Have Internet access at home	Number of computers in home
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

As to be expected, students who reported that they had a computer in the home performed better on CBAS than students who reported that they did not have a computer (although this comparison group was very small). The difference in performance in each country was approximately equal to half a standard deviation as displayed in Figure 16 below. This may be partially explained by socio-economic status as the students who do not possess a computer at home display much lower values on the PISA scale of economic, social and cultural status.

Figure 16.
Performance on CBAS and possession of computer in the home per country



ICT FAMILIARITY ITEMS AND SCALES

All countries participating in CBAS also administered the PISA ICT questionnaire during the PISA paper-and-pencil questionnaire session (along with 37 other countries which contribute to the calculation of the scale indices). This questionnaire asked students a series of 32 questions about the frequency of their computer use for specific activities and their confidence in performing specific activities on the computer. Four scales indices were computed from the collection of items measuring the four areas listed below:

ICT Internet/entertainment use

The index of ICT Internet/entertainment use was derived from students' responses about the frequency with which they use computers for the following reasons: *i*) browse the Internet for information about people, things, or ideas; *ii*) play games; *iii*) use the Internet to collaborate with a group or team; *iv*) download software from the Internet (including games); and *v*) download music from the Internet and *vi*) for communication (*e.g.* e-mail or "chat rooms"). A five-point scale with the response categories "almost every day", "once or twice a week", "a few times a month", "once a month or less" and "never" was used. All items were inverted and positive values on this index indicate high frequencies of ICT use.

ICT program/software use

The index of ICT program/software use was derived from students' responses about the frequency with which they use computers for the following reasons: *i*) write documents (*e.g.* with Word® or WordPerfect®); *ii*) use spreadsheets (*e.g.* Lotus 1 2 3® or Microsoft Excel®); *iii*) drawing, painting or using graphics programs; *iv*) use educational software such as mathematics programs; and *v*) writing computer programs. A five-point scale with the response categories "almost every day", "once or twice a week", "a few times a month", "once a month or less" and "never" was used. All items were inverted and positive values on this index indicate high frequencies of ICT use.

Self-confidence in ICT Internet tasks

The index of self-confidence in ICT Internet tasks was derived from students' beliefs about their ability to perform the following tasks on a computer: *i*) chat online; *ii*) search the Internet for information; *iii*) download files or programs from the Internet; *iv*) attach a file to an e-mail message; *v*) download music from the Internet; and *vi*) write and send e-mails. A four-point scale with the response categories "I can do this very well by myself", "I can do this with help from someone", "I know what this means but I cannot do it" and "I don't know what this means" was used. All items were inverted for IRT scaling and positive scores on this index indicate high self-confidence.

Self-confidence in ICT high-level tasks

The index of self-confidence in ICT Internet tasks was derived from students' beliefs about their ability to perform the following tasks on a computer: *i*) chat online; *ii*) search the Internet for information; *iii*) download files or programs from the Internet; *iv*) attach a file to an e-mail message; *v*) download music from the Internet; and *vi*) write and send e-mails. A four-point scale with the response categories "I can do this very well by myself", "I can do this with help from someone", "I know what this means but I cannot do it" and "I don't know what this means" was used. All items were inverted for IRT scaling and positive scores on this index indicate high self-confidence.

Table 9 below shows the model fit for a four-dimensional model for the ICT familiarity items in PISA 2006. Fit indices measure the extent to which a model, based on a particular structure hypothesised by the researcher, 'fits the data'. Model fit is assessed using Root-Mean Square Error of Approximation (RMSEA),

the Root Mean Square Residual (RMR), the Comparative Fit Index (CFI) and the Non-Normed Fit Index (NNFI). The PISA 2006 Technical Report should be consulted for further information about these techniques (OECD, 2009). Overall, the model fit was considered satisfactory for all of the CBAS participating countries and for the pooled OECD sample.

Table 9.
Model fit for CFA with ICT familiarity items

	Model fit			
Country*	RMSEA	RMR	CFI	NNFI
Denmark	0.099	0.084	0.69	0.70
Iceland	0.089	0.078	0.71	0.72
Korea	0.077	0.060	0.79	0.80
OECD	0.084	0.082	0.81	0.81

* Model estimates based on international student calibration sample (500 students per OECD country).

Table 10 shows the estimated latent correlations for the four ICT familiarity scales. All four constructs are positively correlated with each other; the highest correlations are found between the two constructs reflecting self-confidence in ICT tasks and use of the Internet and programs. Most correlations in Korea are weaker than in other countries, indicating that these scales may not be as good a measure of underlying constructs in Korea as they are in Denmark or Iceland.

Table 10.
Estimated latent correlations for ICT familiarity scales

	Latent correlations between					
	Internet use/	Internet use/	Internet use/	Program use/	Program use/	Internet confidence/
Country	Program use	Internet confidence	High confidence Internet	Internet confidence	High confidence Internet	High confidence Internet
Denmark	0.58	0.42	0.42	0.17	0.42	0.53
Iceland	0.54	0.42	0.41	0.19	0.38	0.57
Korea	0.45	0.19	0.25	0.10	0.45	0.40

Table 11 shows the scale reliabilities for the ICT scales in CBAS countries and the overall median for all PISA countries that administered the ICT familiarity questionnaire. The internal consistencies were mostly high across all PISA countries but were below the median for all CBAS countries, particularly for Internet/ Entertainment and Program/Software use. This should be noted when interpreting the relationship between the scales for frequency of computer use and achievement later in the report.

Table 11.
Scale reliabilities for ICT familiarity scales

Country	Internet use	Program use	Internet confidence	High confidence
Denmark	0.66	0.73	0.76	0.80
Iceland	0.69	0.75	0.74	0.77
Korea	0.66	0.71	0.81	0.82
Median	0.82	0.78	0.87	0.85

ICT FREQUENCY OF USE AND CONFIDENCE ACROSS GENDERS

Female and male scores on the ICT confidence and frequency of use scales were examined across countries and a number of differences were found between the frequency of use of computers and confidence in performing specific tasks. These patterns refer to general differences and have not been tested for statistical significance unless otherwise mentioned.

Males appear to be more frequent users of ICT programs and software than females as there are more males in the upper half of the scale and more females in the lower half. This gender difference in frequency of use is particularly strong in Iceland. Similarly, more females score in the lower half of the ICT Internet and Entertainment Use scale, particularly in Korea and Iceland. Across all three countries, males are more confident than females about performing ICT Internet activities. Further, there are more males who are confident in performing high-level ICT activities, particularly so in Iceland where the majority of males fits in the higher end of the confidence scale.

Frequency of performing specific ICT activities

The following analyses relate to differences observed in the frequency of performing specific activities examined on the ICT questionnaire and not to the ICT scales in general. Actual frequencies for each of the items on the ICT questionnaire by country and by gender are given in Annex B.

Overall, the vast majority of students, both males and females have used computers (greater than 96% across all countries). The majority of students in all three countries have been using computers for five years or more, however, in Iceland more females than males have only been using the internet for one to five years. Icelandic students also browse the Internet more often than students in Korea or Denmark with most students reporting that they browse the Internet almost daily. More males browse the Internet almost every day than females in all three countries.

Yet other activities do not display any gender differences in frequency of usage. For example, there are few gender differences in the frequency of use of the computer for writing documents across all three countries, nor for using educational software. These are relatively infrequent activities.

Male activities on the Internet differ slightly from female activities and the place of use differs slightly in that they are more likely to use computers in other places than home or school than females. Males in all three countries use the computer to play games much more often than females with a majority of males responding that they play games on a computer at least once or twice a week. In Korea and Denmark the most common response for females was that they never play games on the computer and in Iceland female responses were evenly split across once or twice a week, a few times a month, once a month or less, and never. Males in all three countries are much more likely to download music from the Internet than females. Particularly in Iceland, males tend to collaborate more on the Internet than females. Similar to the pattern

observed for playing games on the internet, males in all three countries download software more often than females and they tend to use spreadsheets more than females, although this activity is relatively uncommon across all countries. Similarly, although writing programs is another relatively infrequent activity for PISA students, it is more common for males to perform this activity than females.

Much less numerous are the activities that females tend to perform more frequently than males. They can be considered as 'social networking' activities. Females in Iceland and Denmark tend to use the computer to send and receive emails or visit chat rooms more often than males, (but the vast majority of students in both of these countries report that they use the computer almost every day for these activities). In Korea, there are few gender differences and the majority of students report that they use email or chat rooms almost every day or once or twice a week. The pattern is almost reversed for the downloading of graphics programs in Korea where more females download graphics programs almost daily and more males never do this activity.

Confidence in performing specific ICT activities

Overall, males reported higher confidence for performing computer-based activities than females. This may be as a result of their higher frequency of performing these tasks, however it may also reflect a more general trend observed in that males report higher confidence than females regardless of the activity, competency or familiarity. However there are a couple of commonly performed items that students display equal confidence for regardless of gender; most students, males and females, in all three countries felt confident about chatting via Internet and attaching files to emails.

Females report higher confidence in several items from the questionnaire across countries. For example, there is a slight gender difference in favour of females for editing photos and females appear more confident in making presentations than males. Females are also more confident about sending emails than males in Iceland and Denmark, perhaps reflecting their more frequent use of these activities, although the majority of students in all three countries are confident about sending emails without any help. In Korea there do not appear to be any gender differences.

In Iceland, in contrast to the other two countries, more females than males report high confidence about searching the Internet. Again in Iceland, there is a strong gender difference in the level of confidence about making web pages with many more females displaying higher confidence about making web pages than males. The majority of Icelandic students display higher confidence in making web pages without help, whereas in Denmark and Korea the most common response is that students can do this activity with help and there are no gender differences in degree of confidence. Finally, females are slightly more confident about using a word processor than males in Denmark and Iceland whereas there appears to be no gender difference in Korea.

On the whole, males display greater confidence for advanced activities, particularly activities on the Internet. More males than females feel confident about dealing with a virus on the computer, with the most common response for females in Denmark and Iceland being that they know how to deal with a virus, but they can't do it. Males are more confident about using databases and about copying data to CD. They are also slightly more confident about moving files, though most students say that they can do this well without help. Males are more confident than females about downloading files from the Internet, downloading music and about using spreadsheets which is not surprising given their more frequent use of these three activities. Significantly more males are confident about using multimedia in Denmark and Iceland without help whereas the most common response for females is that they can do these activities with help. In Korea there are no gender differences for this item.

Comparison of male and female confidence and frequency of use

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. As presented in Table 12, overall, and in every country, a higher proportion of males (60%) reported that they were frequent users of ICT (combining program/software use and Internet/Entertainment use) than females (40%).[1]

Table 12.

Proportion of females and males reporting frequent or less frequent use of ICT across countries

	Females		Males	
Country	ICT infrequent users	ICT frequent users	ICT infrequent users	ICT frequent users
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%
TOTAL	59.8%	40.2%	40.3%	59.7%

A similar pattern for confidence in performing basic and advanced ICT tasks is revealed in Table 13 where we can observe that in Denmark and Iceland a majority of males were classified as highly confident whereas a majority of females were classified as having low confidence.[2] In Korea, there were no differences in patterns of confidence between females and males, however one should remember that this scale is highly skewed for Korea. What is notable in these patterns is that although over 60% of Icelandic females report that they are frequent users of the Internet and programs, only 33% report that they feel confident about performing ICT tasks.

Table 13.

Proportion of females and males reporting high or low confidence in performing ICT tasks across countries

	Females		Males	
	ICT low confidence	ICT high confidence	ICT low confidence	ICT high confidence
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%
TOTAL	60.0%	40.0%	40.8%	59.2%

ASSOCIATION BETWEEN ICT FAMILIARITY, CONFIDENCE AND CBAS ACHIEVEMENT

Impact of frequency of ICT use on science achievement

By comparing the usage levels based on the PISA 2006 ICT indices to performance, the extent to which students using computers more across a range of ICT functions tend to do better or worse in the PISA assessment can be analysed. The answers cannot show whether certain kinds of ICT use helps students to perform better at school, but does indicate the extent to which those students who do well are also those students who use ICT for certain purposes.

This analysis looks at two broad indices of usage, one based on how often students use the Internet and play computer games, and the other based on how much they use various computer programs and educational software. Students in each country are divided into four equal groups according to their scores on each index. Those in the highest usage group are those who frequently use computers for a relatively wide range of purposes; those in the lowest are the least frequent users.

Figure 17 and Figure 18 show the average CBAS and paper-and-pencil science scores across countries and genders in relation to the frequency of Internet use. There is a slight tendency towards the U-shaped curve that was observed in the PISA 2003 ICT report for reading and mathematics (OECD, 2005b), where students using the Internet the least scored below those around the middle of the computer use distribution. In the present data, however, it is the students who use the Internet the least frequently who do not support this trend. On the whole, students in this sample who use the Internet the least score higher than students who use the Internet the most and, for the most part, than students who are moderate users of the Internet. Particularly for Danish females, one can notice a substantial drop off in achievement between females in the third quarter and females in the most frequent category of use.

Figure 17.
Students' use of the Internet and average performance in CBAS by quarter of the indices

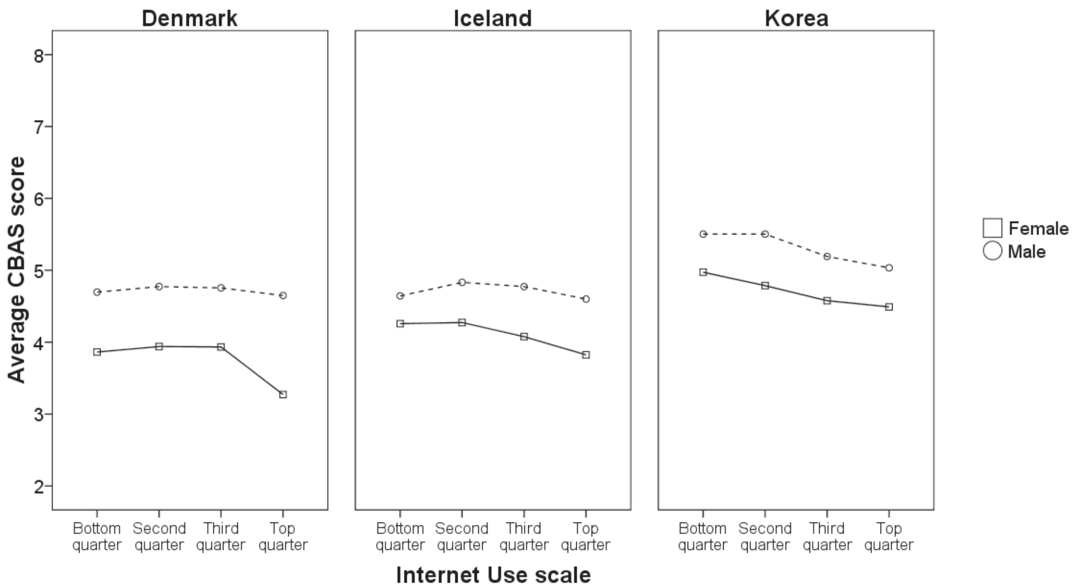
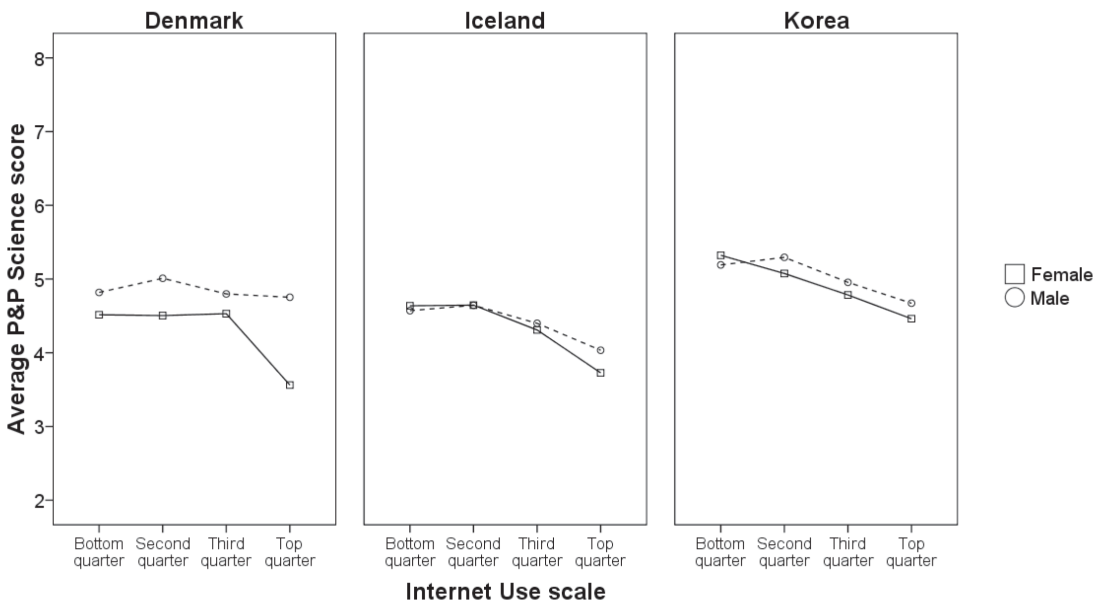


Figure 18.
Students' use of the Internet and average performance in paper-and-pencil science by quarter of the indices

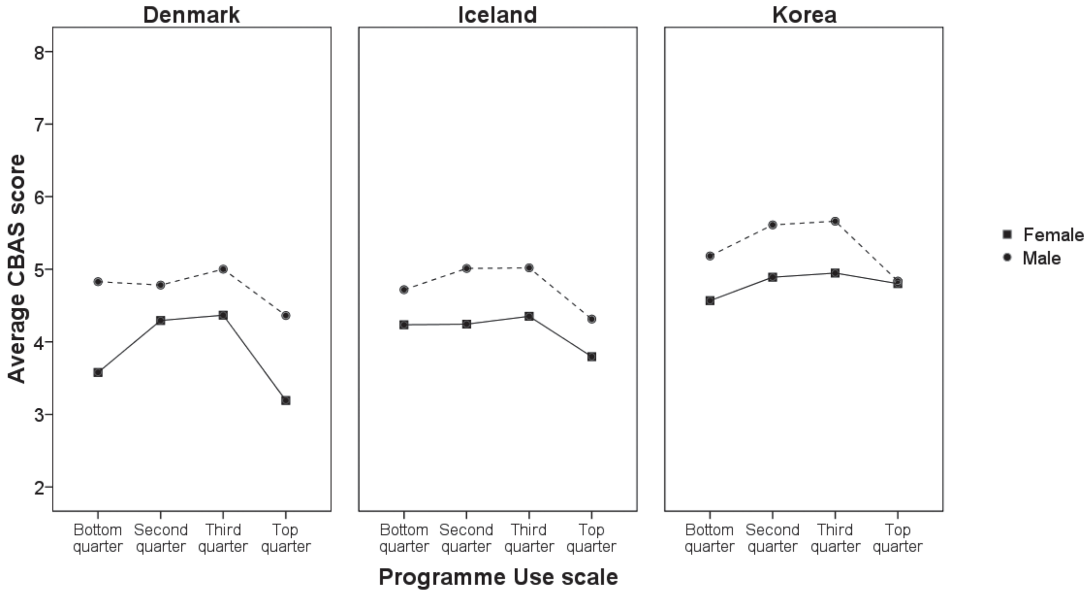


The tendency towards a U-shape is stronger in Figure 19 and Figure 20 below and are therefore better support for the pattern observed in the PISA 2003 ICT report. Overall, students who spend the least and the most time using programs are the lowest achievers and the students who show moderate frequencies of use achieve the highest performances.

For both the paper-and-pencil and the computer-based test there are no clear gender differences across countries. In Denmark it is interesting to note that females in the middle categories almost 'catch up' to males for both paper-and-pencil science and CBAS. In Korea it is the females who use programs the most frequently that catch up on males or overtake them as males' performance drops in this category. In Iceland the gender difference with males performing better than females in CBAS but not in paper-and-pencil science seems to prevail across all categories of usage.

These results reinforce the message that one cannot readily assume that more computer usage is bound to be beneficial for students in all cases. One potential explanation of these results is that the students reporting the highest use of the Internet, playing games and using programs are doing so to the detriment of homework and other out-of-school learning activities.

Figure 19.
Students' use of programs and educational software and average performance in CBAS by quarter of the indices



Impact of ICT confidence on science achievement

By comparing the confidence levels based on the PISA 2006 ICT indices to performance, the extent to which students who are more or less confident in using computers across a range of ICT functions, tend to do better or worse in the PISA assessment can be analysed. The answers cannot show whether confidence in ICT use helps students to perform better at school, but do indicate the extent to which those students who do well are also those students who are confident performing different ICT tasks.

On the whole, the patterns of confidence displayed by females and males in Figure 21 through 24 are similar. In Denmark the trend towards a U-shaped curve appears to be the strongest out of all countries, particularly for the females. Note that in Iceland, the trend towards a U-shaped curve is stronger for achievement on the paper-and-pencil test than on the computer-based test, but on the whole the distributions appear relatively flat, indicating only a weak association between confidence and achievement.

Figure 20.

Students' use of programs and educational software and average performance in paper-and-pencil science by quarter of the indices

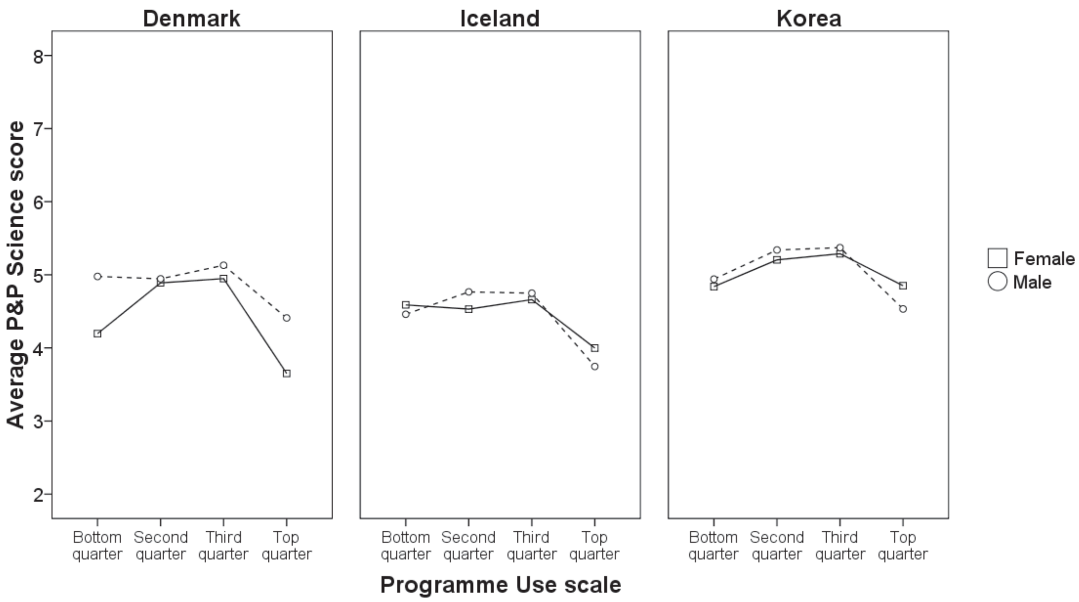


Figure 21.

Internet confidence and CBAS achievement across countries and genders[3]

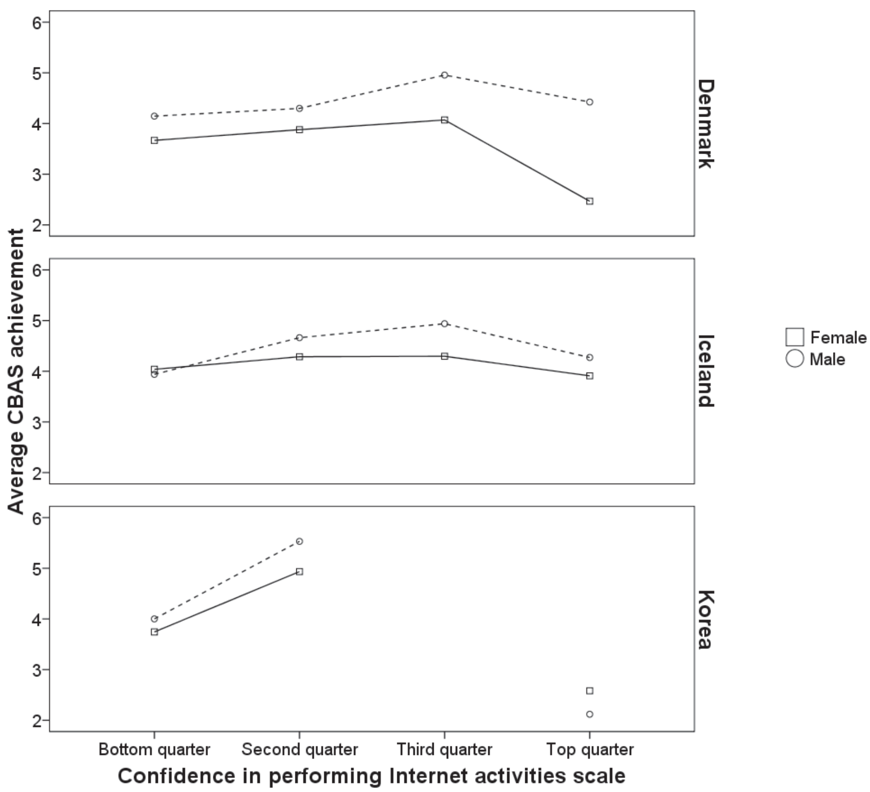


Figure 22.
Internet confidence and paper-and-pencil science achievement across countries and genders

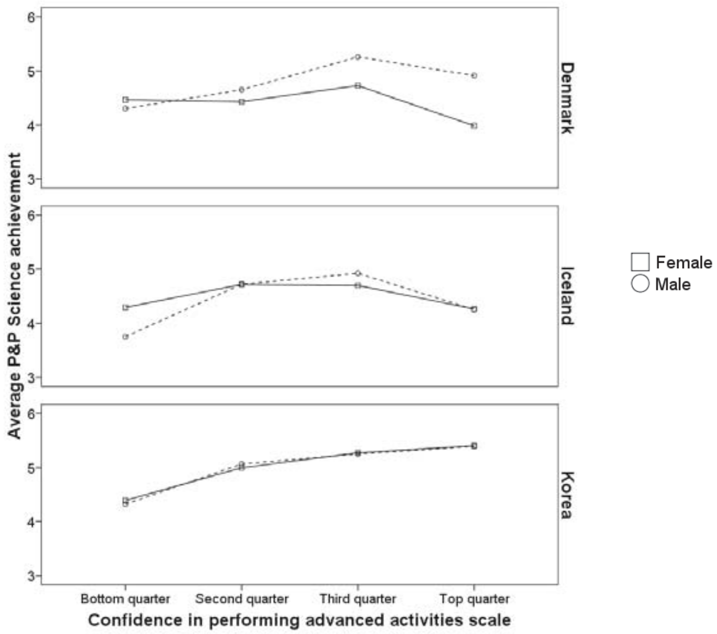


Figure 23.
Advanced ICT activities confidence and CBAS achievement across countries and genders

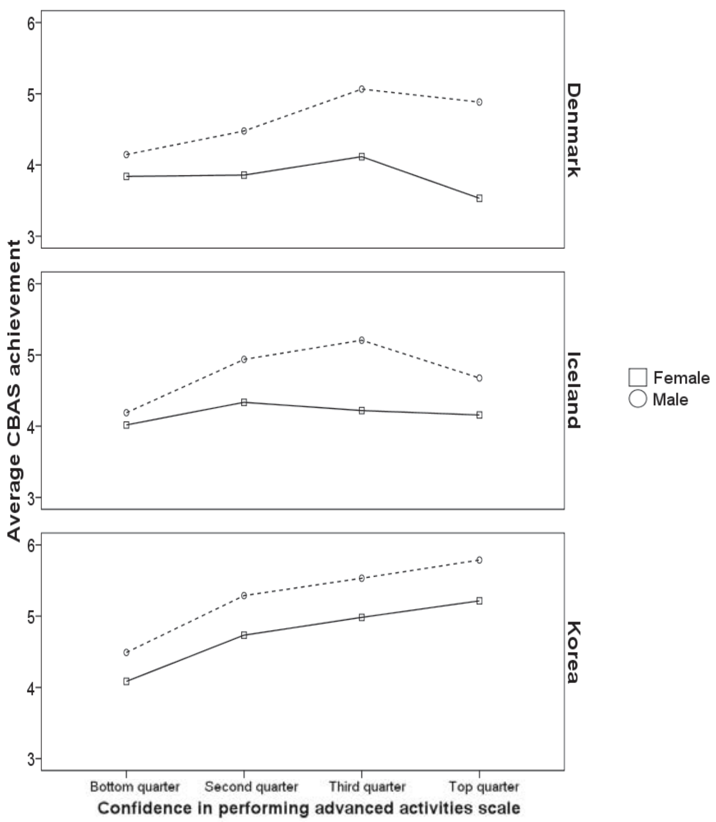


Figure 24.
Advanced ICT activities confidence and paper-and-pencil science achievement across countries and genders

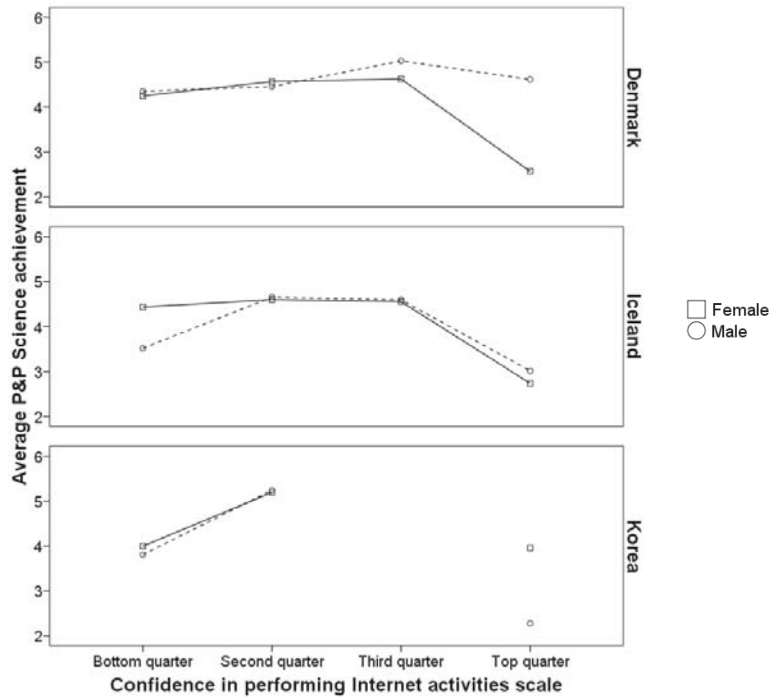
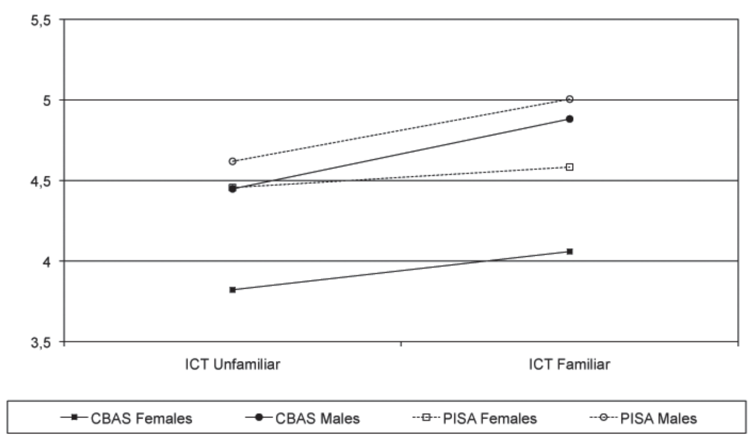


Figure 25.
Impact of ICT familiarity for male and female students on CBAS and paper-and-pencil scores in Denmark



ICT familiarity and CBAS achievement per country

The following analyses investigate the relationship between the combined measure of ICT familiarity and confidence on males' and females' performance in CBAS and paper-and-pencil science for each country.

Figure 25 shows that in Denmark, ICT familiar females and males performed better than their 'ICT-unfamiliar' counterparts on both the CBAS and paper-and-pencil test, although the size of the difference between familiar and unfamiliar students was smaller for females. Additionally, the data show that ICT familiarity is positively correlated with performance for males and females across both test modalities. This effect may be partially explained by socio-economic status which is positively correlated with ICT familiarity[4] in the PISA paper-and-pencil test, as shown in Table 14 below, socio-economic background was positively associated with a student's performance and the strength of the relationship was approximately similar in all three CBAS countries.

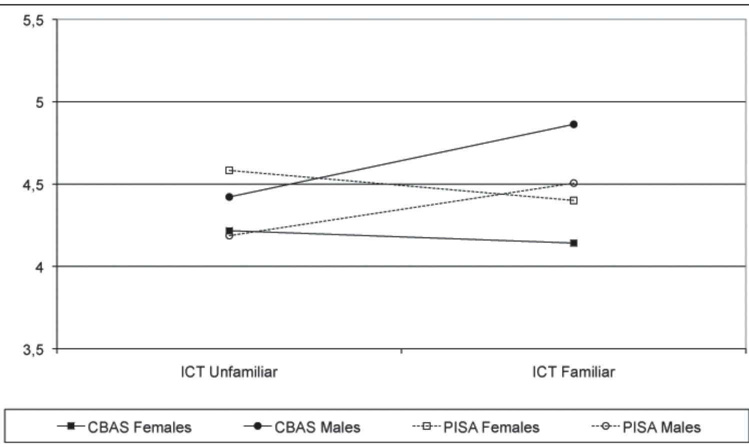
Table 14.
Score point difference on science scale associated with one unit on the ESCS (gradient)*

	Score point difference
Denmark	0.78
Iceland	0.58
Korea	0.64

*Adjusted to a similar scale used for CBAS

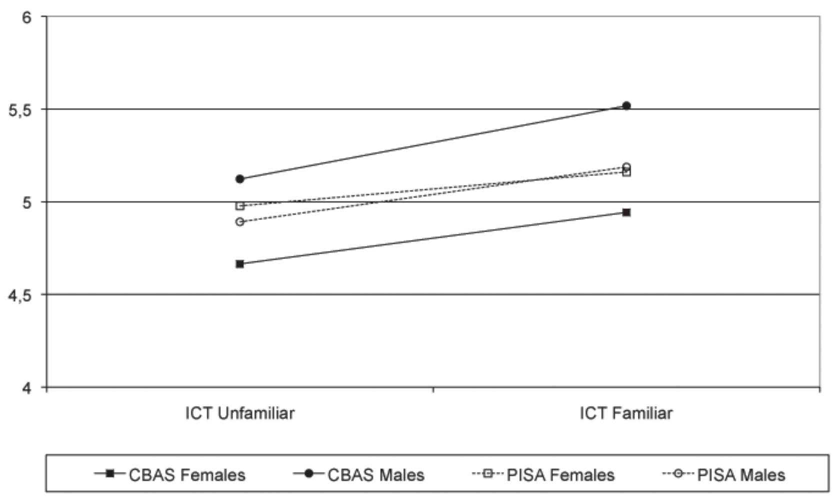
Figure 26 reveals that in Iceland the same pattern is present for the males, but the reverse pattern is observed for the females: high ICT familiarity is associated with poorer performance for females on both the CBAS and the paper-and-pencil tests. On the paper-and-pencil test, ICT unfamiliar females outperformed their ICT unfamiliar male counterparts. Nevertheless, the reverse was true for the ICT familiar students where males outperformed females on the paper-and-pencil test of science and on CBAS. The Icelandic females are the only group out of the three countries to display a negative correlation between ICT familiarity and achievement. This may reflect the types of activities that Icelandic females are performing on computers if these activities are not educational and time spent on the computer replaces other educational activities such as homework or out-of-school lessons. This pattern requires further investigation in the future to identify what sorts of females are ICT unfamiliar and why their performance on both the computer-based and the paper-and-pencil test is disadvantaged.

Figure 26.
Impact of ICT familiarity for male and female students on CBAS and paper-and-pencil scores in Iceland



As shown in Figure 27, in Korea, ICT familiarity is again associated with higher scores on both CBAS and PISA for both genders, however this effect is stronger on the CBAS test than on the paper-and-pencil test and stronger for males than for females. Further, while males who are ICT familiar did not outperform females who are ICT familiar on the paper-and-pencil test, the same males do perform better than females who are ICT familiar on the CBAS test. This suggests that the CBAS test favours ICT familiar males, however it should be kept in mind when interpreting these results that the gender difference between males and females overall in Korea was not statistically significant.

Figure 27.
Impact of ICT familiarity for male and female students on CBAS and paper-and-pencil scores in Korea



These differences are summarised in the following Table 15 showing the size of the advantage for ICT familiar females and males in comparison to ICT unfamiliar students and whether these differences were significant or not. The advantage for ICT familiar males over ICT unfamiliar males is almost ¼ of a standard deviation, whereas for females the only significant advantage is for females in Korea and here the advantage is smaller. The Icelandic females stand out here once again since they are the only group for whom there is no trend towards a performance advantage for ICT familiar students (although the advantage for females in Denmark is also not significant due to the large standard error, a definite trend in this direction is present).

Table 15.
Effects of ICT familiarity on performance for male and female students across countries

	Advantage for ICT familiar female	S.E.	Advantage for ICT familiar male	S.E.
Denmark	0.21	0.21	**0.46**	**0.19**
Iceland	0.07	0.10	**0.38**	**0.16**
Korea	**0.29**	**0.13**	**0.43**	**0.16**

*Significant differences are displayed in bold

Table 16 below presents this relationship in another way, displaying the correlations for familiarity and achievement across countries which are stronger for males than for females (although on the whole they are relatively weak).

Table 16.
Correlations between ICT familiarity science scores on the CBAS and paper-and-pencil tests

	Paper-and-pencil		CBAS	
	Female	Male	Female	Male
Denmark	0.03	0.09	0.06	0.11
Iceland	-0.04	0.07	-0.02	0.10
Korea	0.05	0.07	0.07	0.10

Notes

1. 2x2 Chi Square results. Denmark ($\chi^2(1,1158=70) = 0.0$, $p<0.001$), Iceland ($\chi^2(1,3400=227) = 0.0$, $p<0.001$) and Korea ($\chi^2(1,2626=16) = 0.0$, $p<0.001$)

2. 2x2 Chi Square results: Denmark ($\chi^2(1,1154 =98) = 0.000$, $p<0.001$), Iceland ($\chi^2(1,3273=309) = 0.000$, $p<0.001$) and Korea ($\chi^2(1,2412 =0.9) = 0.182$, $p>0.05$)

3. ESCS – the PISA index of economic, social and cultural status that summarises various aspects of socio-economic background, including the occupational status and level of education of the students' father and mother and students' access to educational and cultural resources at home – and ICT familiarity are moderately correlated at 0.31

4. CBAS: Denmark (FET =6, $p>0.05$), Iceland (FET =16, $p<0.05$) and Korea (FET =12, $p<0.05$)

Chapter 5

5

CBAS questionnaire results

Male responses to questionnaire items were more polarised than female replies. They usually strongly disagreed or strongly agreed with statements more than females, who tended towards more neutral 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 prefered 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 seemed to have little to do with achievement. Test preference and relative effort reports showed no association with test performance.

The CBAS questionnaire was administered so that the relationship between achievement and test engagement factors (enjoyment, motivation and effort) could be investigated. This section examines firstly whether there were any differences across countries and genders in terms of test engagement and secondly, whether these differences (if any) can explain variations in performance between tests.

Male and female responses to the questions regarding enjoyment and motivation for the computer-based and paper-and-pencil tests are displayed in Table 17 below. Each question is then discussed individually in the following sections.

Table 17.
Male and female students responses to the CBAS questionnaire across countries

			Strongly Disagree		Disagree		Agree		Strongly Agree	
			n	%	n	%	n	%	n	%
"I found the computer-based test enjoyable"	Denmark	Females	19	4,5	37	8,7	243	57,2	126	29,6
		Males	26	6,5	35	8,7	198	49,1	144	35,7
	Iceland	Females	56	15,2	124	33,6	172	46,6	17	4,6
		Males	88	22,9	105	27,3	155	40,4	36	9,4
	Korea	Females	18	2,7	80	12,0	367	55,2	200	30,1
		Males	43	5,4	88	11,2	385	48,8	273	34,6
"I found the paper-and-pencil test enjoyable"	Denmark	Females	61	14,6	172	41,1	168	40,1	18	4,3
		Males	99	25,2	152	38,7	124	31,6	18	4,6
	Iceland	Females	129	35,8	143	39,7	78	21,7	10	2,8
		Males	144	38,8	145	39,1	72	19,4	10	2,7
	Korea	Females	93	14,0	328	49,3	199	29,9	45	6,8
		Males	161	20,4	349	44,2	216	27,4	63	8,0
"I would do another computer-based test just for fun"	Denmark	Females	90	21,2	148	34,8	149	35,1	38	8,9
		Males	100	24,8	150	37,2	95	23,6	58	14,4
	Iceland	Females	143	38,8	128	34,7	77	20,9	21	5,7
		Males	146	38,0	112	29,2	91	23,7	35	9,1
	Korea	Females	165	24,8	329	49,5	139	20,9	32	4,8
		Males	147	18,6	350	44,4	218	27,6	74	9,4
"I would do another paper-and-pencil test just for fun"	Denmark	Females	143	33,9	178	42,2	86	20,4	15	3,6
		Males	180	45,0	165	41,3	38	9,5	17	4,3
	Iceland	Females	162	43,9	128	34,7	69	18,7	10	2,7
		Males	155	40,4	142	37,0	69	18,0	18	4,7
	Korea	Females	147	22,1	276	41,5	207	31,1	35	5,3
		Males	151	19,1	324	41,1	247	31,3	67	8,5

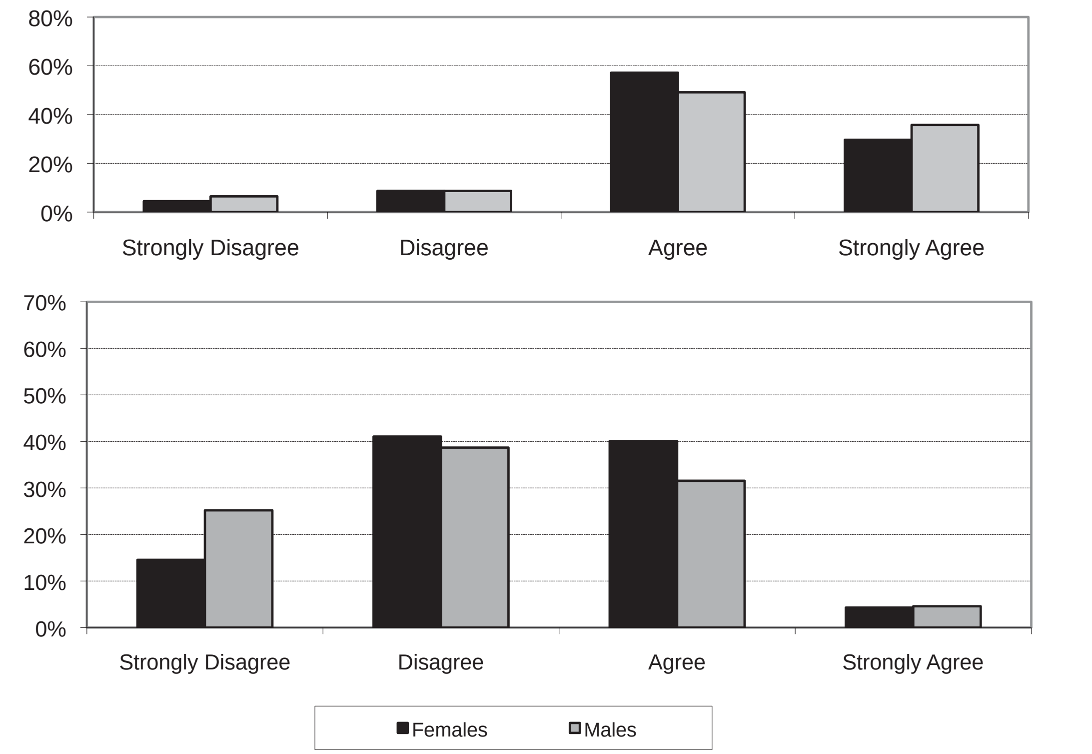
TEST ENJOYMENT ACROSS TEST MODALITIES AND GENDERS

Students were asked to rate on a four-point Likert scale how much they enjoyed the paper-and-pencil and computer-based tests and whether they would do a test "just for fun" if the answers were given at the end. Fisher's exact probability test was performed to investigate whether there were any differences between enjoyment reported by males and females in each country.

Enjoyment of the CBAS test results indicate clearly different patterns of enjoyment for females and males in Iceland and Korea, but no differences in Denmark.[1] On the whole, more males strongly agree and strongly disagree that they enjoyed the test, whereas females tend towards the middle categories of agree and disagree. These patterns are illustrated in Figure 28 to Figure 30 below, which are grouped country-by-country so as to avoid comparisons between culturally-specific response patterns.

Figure 28.

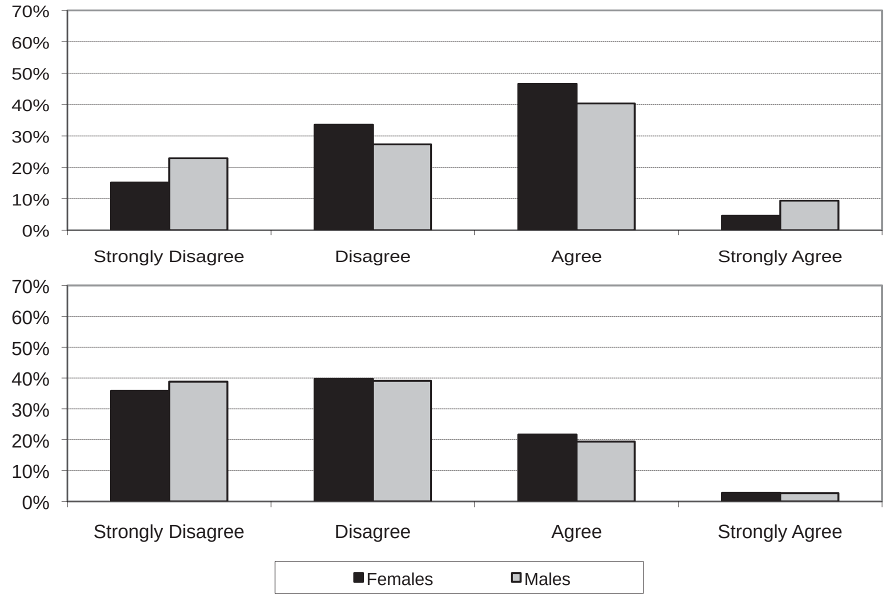
Danish students' endorsement of the statement "I found the CBAS test enjoyable" (top) and endorsement of the statement "I found the paper-and-pencil test enjoyable" (bottom)



On the paper-and-pencil test, the Fisher's exact probability test identifies differences in the patterns of enjoyment for females and males in Denmark and Korea. Here one can see a middle tendency for the females with more females than expected agreeing or disagreeing that they enjoyed the test. More males on the other hand, as in CBAS, tend to strongly disagree and strongly agree that they enjoyed the test. In Iceland, the tendency indicates that females enjoyed the paper-and-pencil test more than males with more females than males strongly agreeing that they enjoyed the test and more males than females strongly disagreeing that they enjoyed the test.[2]

Figure 29.

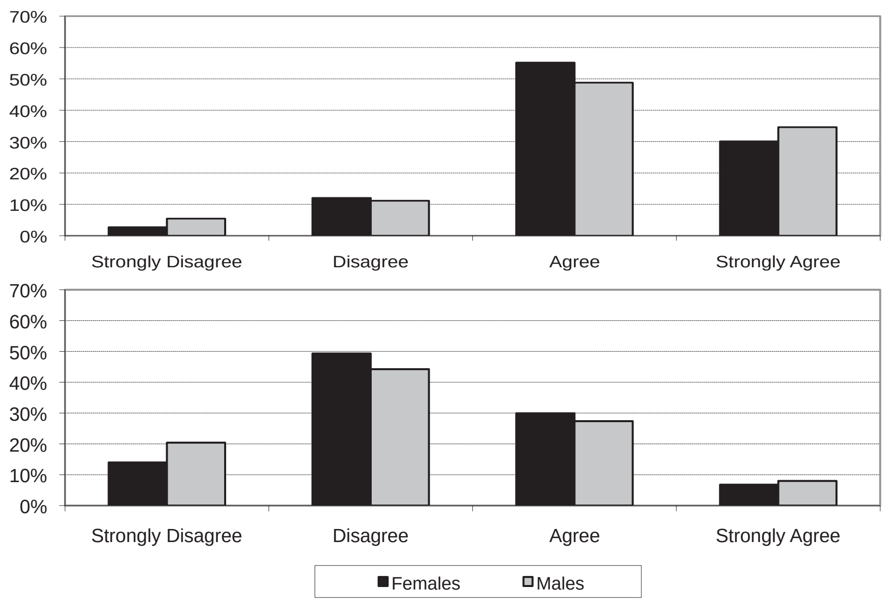
Icelandic students' endorsement of the statement "I found the CBAS test enjoyable" (top) and endorsement of the statement "I found the paper-and-pencil test enjoyable" (bottom)



In comparing Figure 29 to Figure 28 and Figure 30, one can note that Icelandic students overall show less enjoyment of the CBAS and the paper-and-pencil test compared to students in Denmark and Korea, indicating a specifically cultural pattern of low enjoyment reported by students.

Figure 30.

Korean students' endorsement of the statement "I found the CBAS test enjoyable" (top) and endorsement of the statement "I found the paper-and-pencil test enjoyable" (bottom)

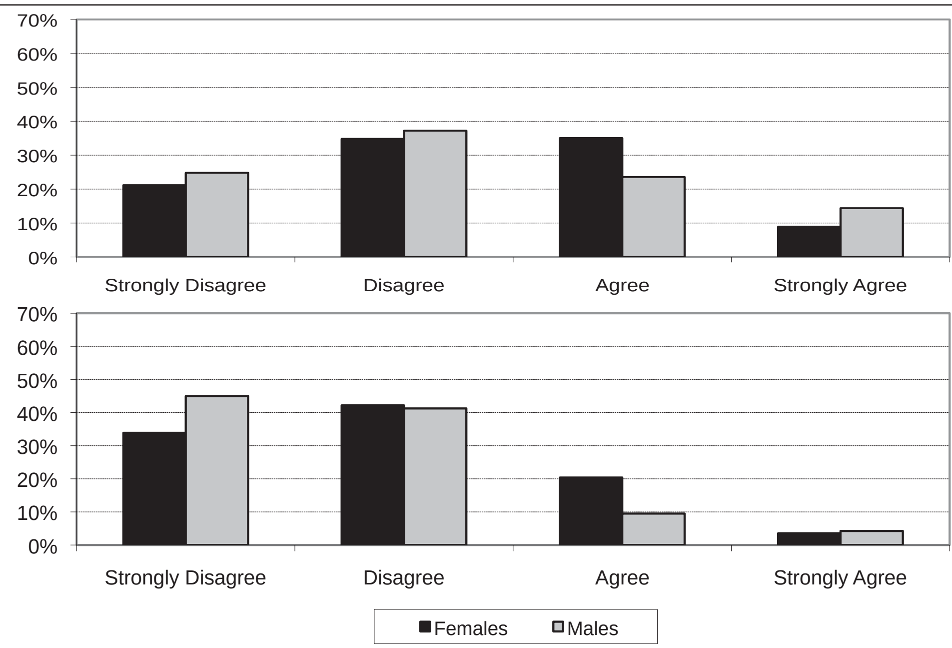


TEST MOTIVATION ACROSS TEST MODALITIES AND GENDERS

Similar to the results displayed regarding test enjoyment, Figure 31 below shows that in Denmark, males responses are more polarised than the female with more males strongly disagreeing with the statement that they would do the computer-based test "just for fun" and more males than females strongly agreeing that they would do the test "just for fun". (Fisher's exact test for probability shows that this difference is statistically significant (FET =16, $p<0.05$). The figure also shows that motivation on the paper-and-pencil test was weaker with a stronger tendency for all students to disagree to 'do the test just for fun'. However, again we observe a middle tendency for the females who are more likely to agree or disagree to do the paper-and-pencil test just for fun than males who are, in return, more likely to strongly disagree or strongly agree. (The Fisher's exact test for probability shows this to be a statistically significant pattern (FET=23, $p<0.05$).

Figure 31.

Danish students' endorsement of the statement "I would do another computer-based test for fun" (top) and endorsement of "I would do another paper-and-pencil test for fun"



The pattern in Figure 32 in Iceland indicates a clearer trend with males more motivated than females on the CBAS test. Females are more likely than males to strongly disagree or disagree to do the computer-based test "just for fun" whereas males are more likely than females to agree or strongly agree. The Fisher's exact test reveals however that these differences are not significant (FET =5, $p>0.05$). Note that the patterns of motivation are relatively similar for the paper-and-pencil test of motivation with most Icelandic students disagreeing or strongly disagreeing to do the test "just for fun". No gender differences are however apparent in motivation for this test (FET =3, $p>0.05$). Again, Icelandic students appear to be the 'least motivated' out of students from all three countries, with the most common response being that they strongly disagree to do another test (regardless of modality) "just for fun".

Figure 32.

Icelandic students' endorsement of the statement "I would do another computer-based test for fun" (top) and endorsement of "I would do another paper-and-pencil test for fun" (bottom)

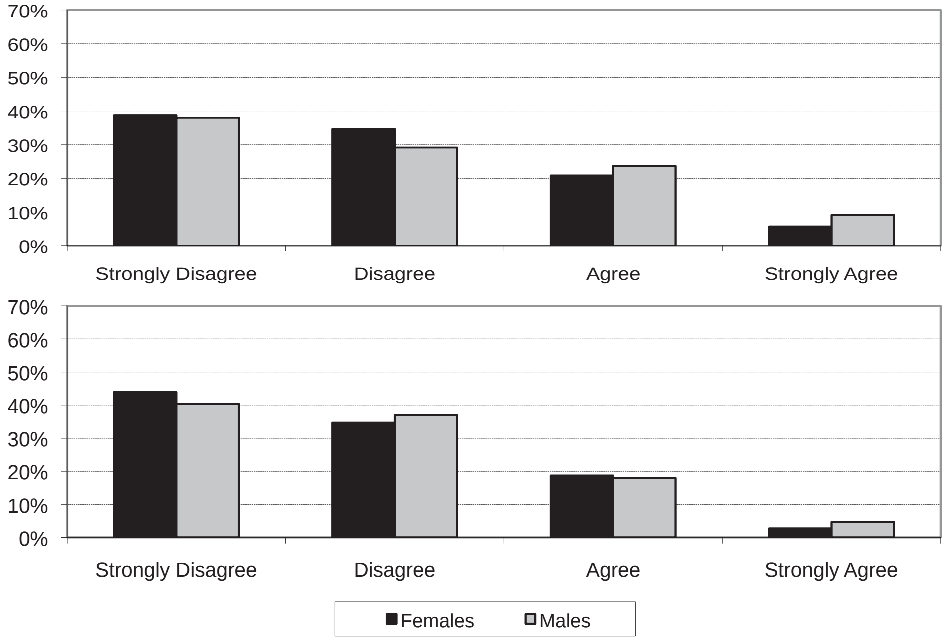
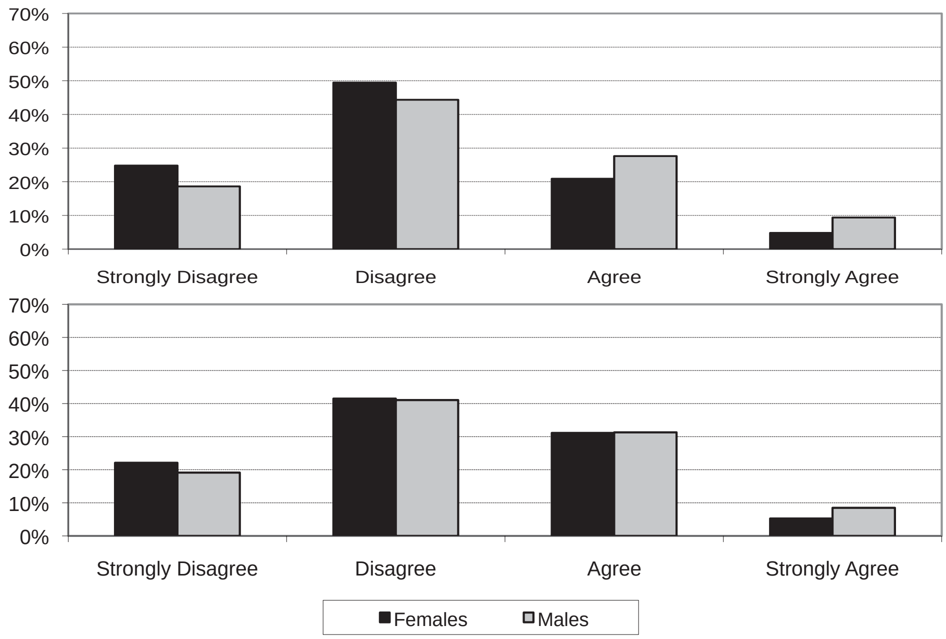


Figure 33.

Korean students' endorsement of the statement "I would do another computer-based test for fun" (top) and endorsement of "I would do another paper-and-pencil test for fun" (bottom)



Similarly, as Figure 33 displays, in Korea the males appear more motivated on the CBAS test with more females than males disagreeing or strongly disagreeing that they would do the computer-based test "just for fun" and more males than females agreeing or strongly agreeing that they would do the test "just for fun". The Fisher's test reveals that this is statistically significant (FET =25, $p<0.05$). It is interesting to note that whereas approximately half of all students in Korea disagree to do the computer-based test just for fun, much fewer would disagree to do the paper-and-pencil test "just for fun". This pattern is unique to Korea and is in direct contrast to the results observed in Denmark and Iceland. In terms of gender difference in motivation on the paper-and-pencil test, there is nothing to indicate that there are gender differences in the degrees of motivation (FET =7, $p>0.05$).

RELATIVE EFFORT APPLIED TO TESTS ACROSS GENDERS

Table 18 displays the actual frequencies of responses to the question "Which test did you try harder on". Gender differences per country are analysed and discussed in the following paragraphs.

Table 18.
Relative effort across genders and tests

		Put more effort into the CBAS test (N)	Same amount of effort on both tests (N)	Put more effort into paper-and-pencil test (N)
Denmark	Females	81	259	81
	Males	50	259	88
	Total	131	518	169
Iceland	Females	57	246	59
	Males	56	235	90
	Total	113	481	149
Korea	Females	123	398	145
	Males	148	449	189
	Total	271	847	334

As the responses to the question 'Which test did you put more effort into' in Figure 34 below show, the majority of students say that they put about equal amounts of effort into both tests. This contrasts with the effort thermometer results that we will examine later that indicate that students tried harder on CBAS, however it should be noted that these results show a strong middle tendency and that the effort thermometer may be a more sensitive measure. Fisher's exact test reveals that a higher proportion of males in all three countries report that they put more effort into the CBAS test, whereas a higher proportion of females in Denmark and Iceland report that they put more effort into the paper-and-pencil test.[3]

Preference for the computer-based or paper-and-pencil test

Table 19 below displays the frequency counts for female and male responses to the question asking students to choose between a two hour paper-and-pencil test, a two hour computer-based test or one hour of each. Gender differences in each country are analysed and discussed in the following paragraphs.

Figure 35 clearly shows that on the whole, more students would prefer to do two hours of computer-based testing than two hours of paper-and-pencil testing. This directly supports the work of Singleton, 2001 and Zandvliet and Farragher, 1997 who found that students preferred computer-based testing to paper-and-pencil testing. In every country there were more females than males who would choose one hour of each test. Fisher's exact probability test results however reveal that this gender difference is only significant in Iceland and Korea, and that in both of these countries there is a strong middle tendency from the females while the males out-represent the females both in the two categories of preference. Here again, it seems that females are more likely to "sit on the fence" or err towards the middle of an opinion-based continuum than males.[4]

Figure 34.
Male and female students' relative effort across assessments

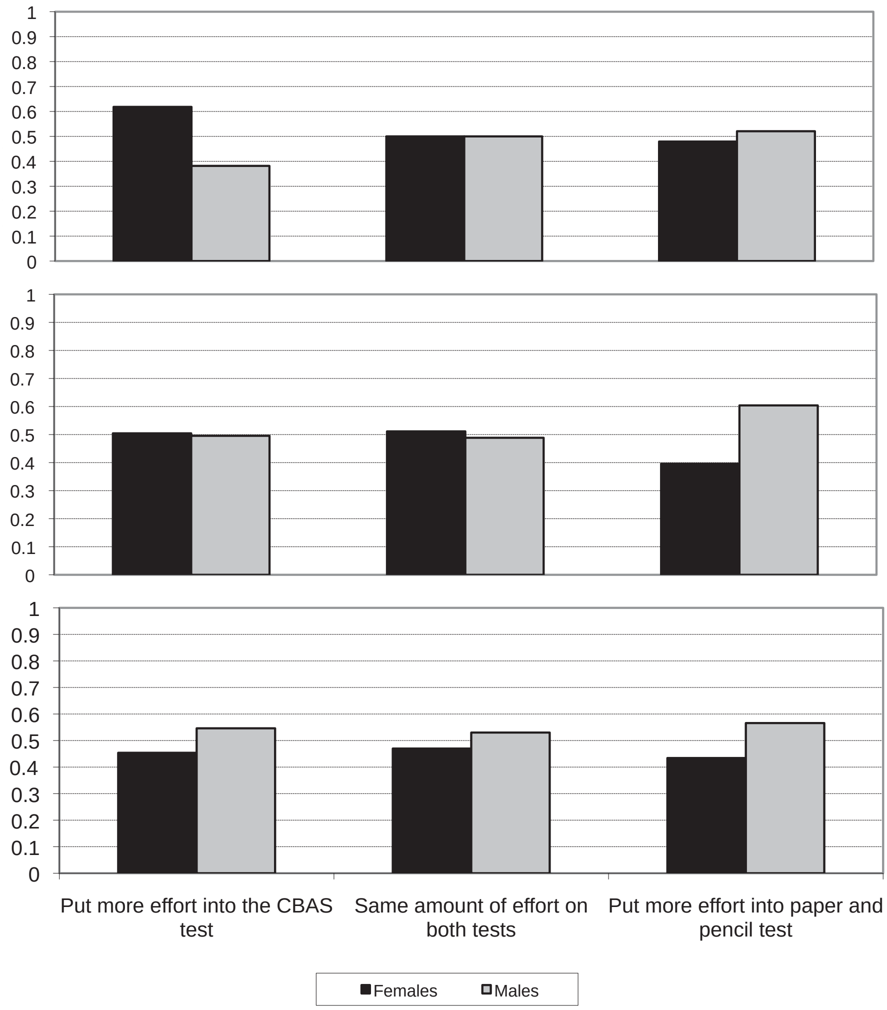
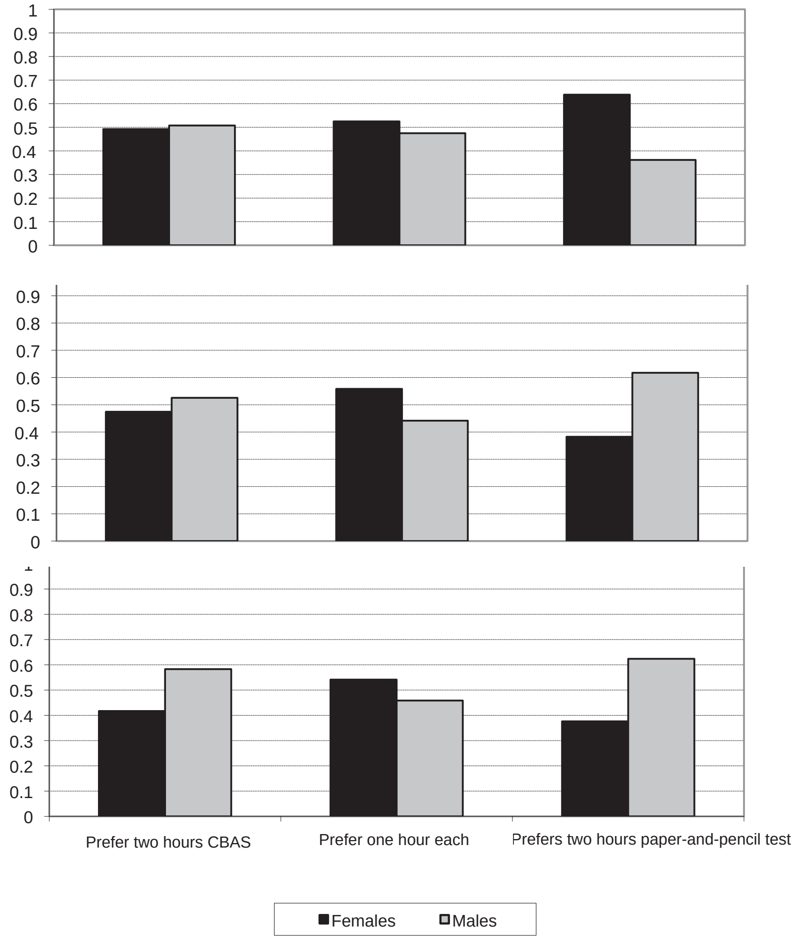


Table 19.
Female and male students' preference for the CBAS or paper-and-pencil test across countries

		Prefer two hours CBAS (N)	One hour each (N)	Prefer two hours paper-and-pencil (N)
Denmark	Females	226	168	30
	Males	233	152	17
	Total	459	320	47
Iceland	Females	167	158	44
	Males	185	125	71
	Total	352	283	115
Korea	Females	347	281	38
	Males	485	238	63
	Total	832	519	101

Figure 35.
Students' preference for the computer-based or paper-and-pencil test



Impacts of motivation and enjoyment on achievement

Figure 36 and Figure 37 show that, overall, there is no association between achievement and test motivation or test enjoyment. The only pattern appearing in Denmark seems to be that as males' motivation increases, their achievement decreases. In Iceland, again the females' response pattern differentiates from all other groups, as they are the only group to show a real association between higher motivation and enjoyment and higher achievement. Males in Iceland also slightly support this pattern, however, achievement drops off in the highest category of motivation. In Korea, greater enjoyment is to some extent associated with higher achievement for males and females.

If the relationship between enjoyment and performance is examined further by considering the correlations in Table 20, one sees that the link between these enjoyment and motivational factors and achievement is in fact very weak or non-existent. The only group that indicates a consistent association between higher motivation and enjoyment and higher achievement are the Icelandic females. The more the Icelandic females students enjoy the test or feel motivated towards the test, the higher they achieve, although the link is not strong. This is strong evidence to indicate that higher scores on the computer-based test, particularly for males, cannot solely be explained by greater enjoyment of the test.

Figure 36.
Motivation for CBAS test and achievement

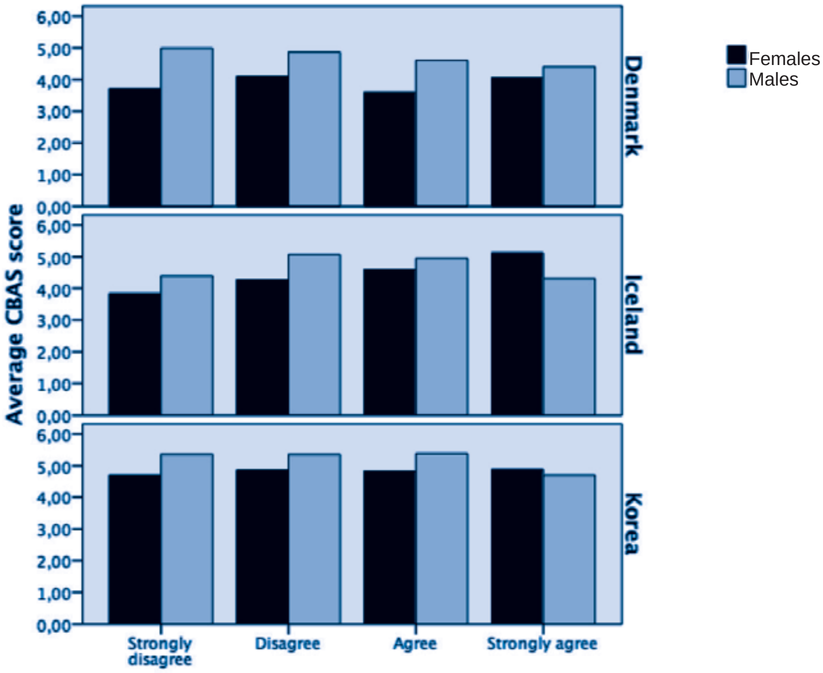


Figure 37.
Enjoyment of CBAS test and achievement

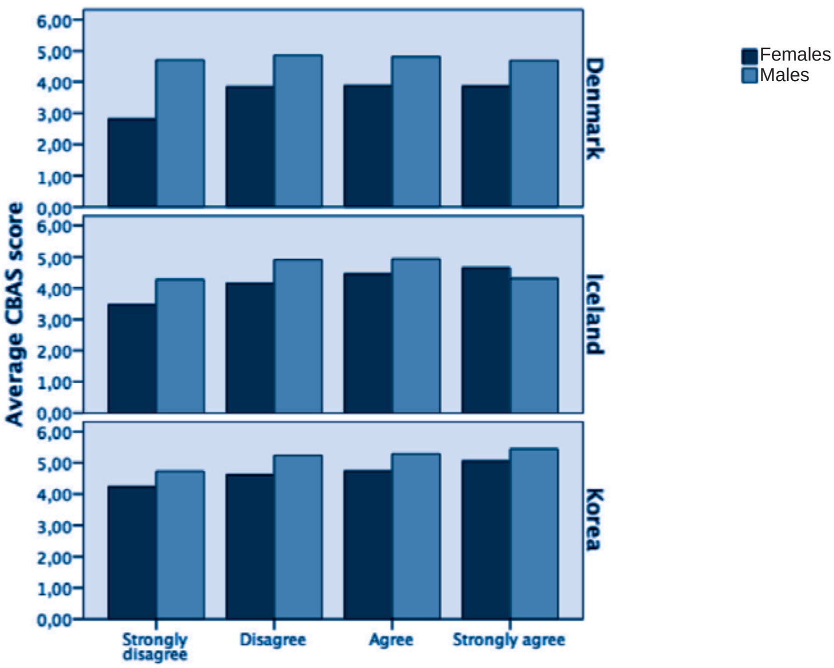


Table 20.

Correlations between CBAS enjoyment, motivation and achievement

	CBAS Enjoyment and achievement		CBAS Motivation and achievement		CBAS Motivation and enjoyment	
	Females	Males	Females	Males	Females	Males
Denmark	0.074	-0.020	-0.011	-0.117	0.347	0.386
Iceland	0.183	0.064	0.201	0.056	0.442	0.507
Korea	0.098	0.073	0.025	-0.058	0.319	0.310

*Correlations significant at the 0.05 level (2-tailed) are displayed in bold.

Figure 38 below displays the relationship between test motivation and the average performance on the paper-and-pencil science scale graphically. Here we see that as motivation increases for females in Denmark and Korea, performance also increases. For Korean males and Icelandic students, performance increases with motivation until the highest category of motivation giving an inverted U-shape distribution. Figure 39 shows the relationship between enjoyment and average achievement indicating that as enjoyment increases in Denmark, so does performance. The effects are similar in Korea but the relationship weaker. In Iceland, we see two inverted U-shape curves with performance increasing for males until the last category and performance increasing for females until the second response category. On the whole, no distinct patterns emerge other than a slightly stronger relationship between motivation and achievement for females, yet the correlations showed this to be weak.

Figure 38.

Motivation for paper-and-pencil test and paper-and-pencil science achievement

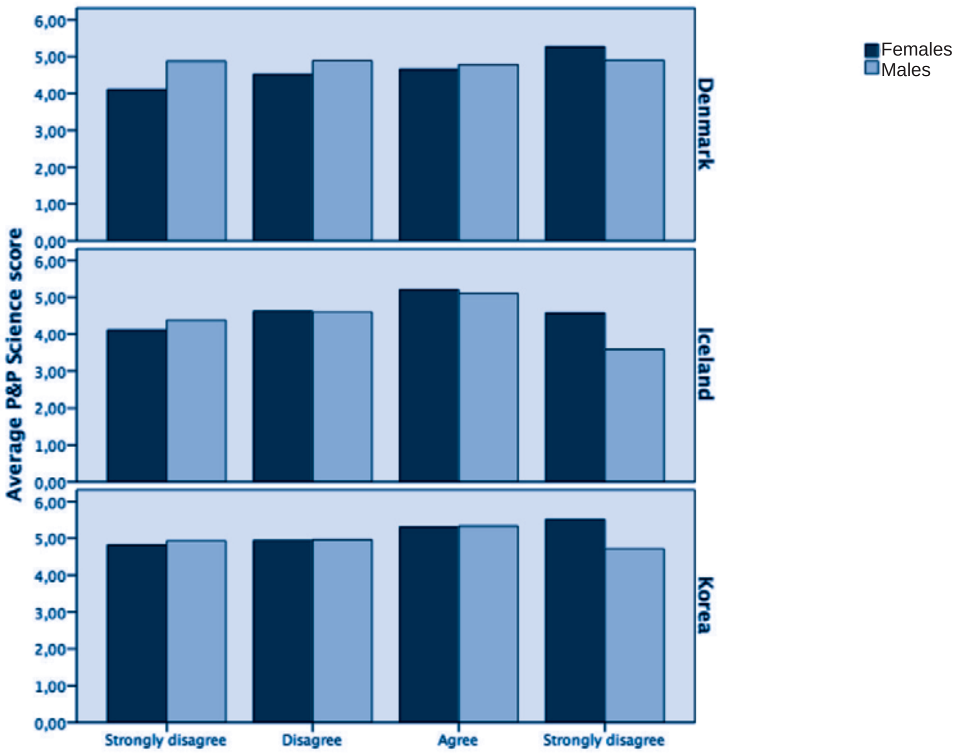


Figure 39.
Enjoyment of paper-and-pencil test and paper-and-pencil science achievement

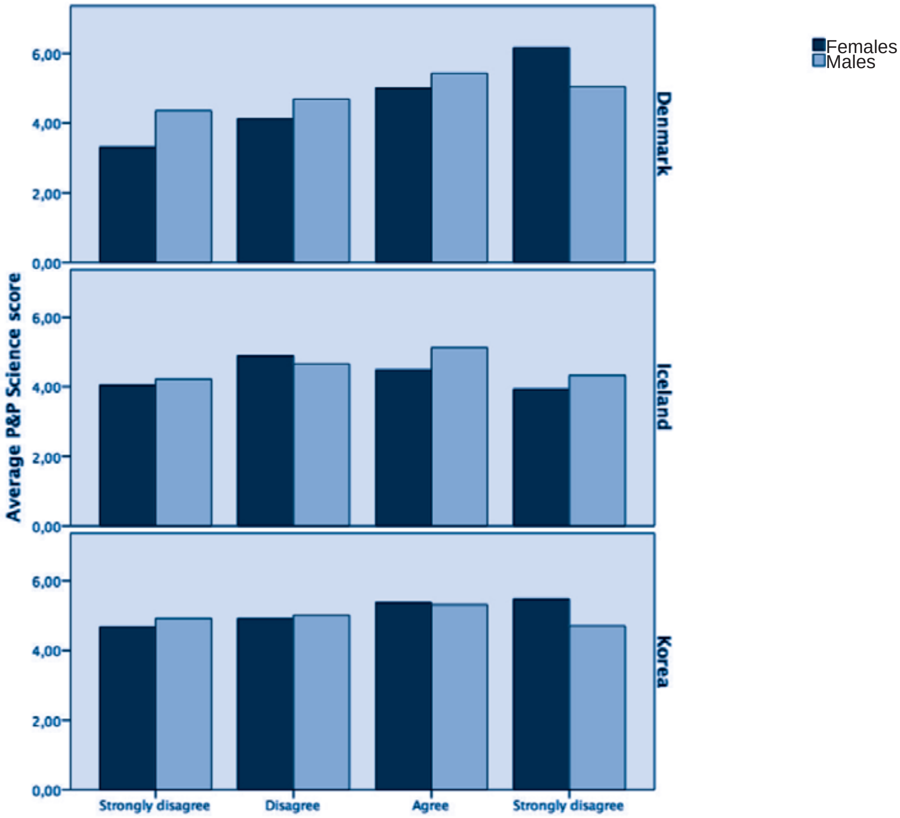


Table 21 shows the correlations between enjoyment, motivation and achievement on the paper-and-pencil test. In Denmark greater enjoyment of the paper-and-pencil test is associated with higher achievement for both males and females. For females in all three countries greater motivation for the test is associated with higher performance although all of these correlations are weak at best. However, this relationship is strongest in Iceland where females outperformed the males on the assessment of science by paper-and-pencil. Overall though, the poor strength of these relationships, along with the lack of clear patterns across countries, indicate that it is clearly not enjoyment and motivation that are causing the increase in performance for males when the test is presented via the computer.

Table 21.
Correlations between paper-and-pencil enjoyment, paper and pencil motivation and paper-and-pencil achievement

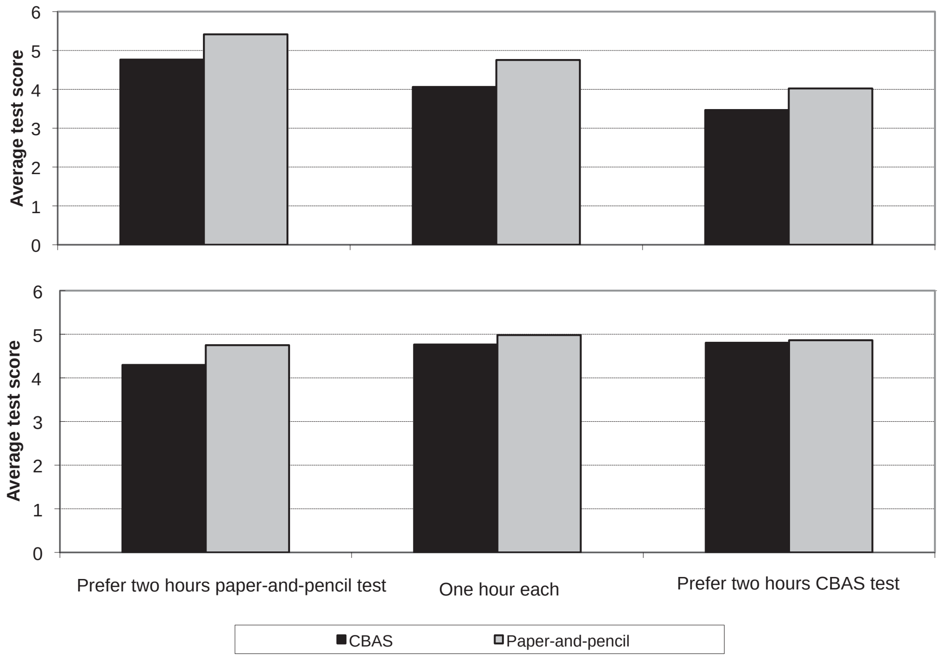
	Paper-and-pencil enjoyment and achievement		Paper-and-pencil motivation and achievement		Paper-and-pencil motivation and enjoyment	
	Females	Males	Females	Males	Females	Males
Denmark	**0.331**	**0.199**	**0.125**	-0.005	**0.432**	**0.418**
Iceland	0.074	**0.135**	**0.179**	0.044	**0.388**	**0.474**
Korea	**0.138**	0.025	**0.112**	0.031	**0.360**	**0.343**

*Correlations significant at the 0.05 level (2-tailed) are displayed in bold.

IMPACTS OF TEST PREFERENCE AND RELATIVE EFFORT ON ACHIEVEMENT

The three figures below display the relationship in each country between test preference and achievement on the paper-based and the computer-based tests. Figure 40 shows that in Denmark the patterns are quite different for females and males. Females that would prefer a two-hour computer-based test score lower than females that prefer the paper-based test and those that would choose equal amounts of both. In contrast, the males who prefer the paper-based test score below the males that prefer the CBAS test and those that would choose equal amounts of both.

Figure 40.
Average CBAS and paper-and-pencil science achievement and test preference for Danish females (top) and males (bottom)



The Icelandic results in Figure 41 do not show any clear differences in achievement based on gender, nor on preference for either type of test.

Figure 42 shows that females in Korea who prefer to do a two hour paper-based test actually score well below females who would choose equal amounts of both types of tests or who would prefer the computer-based test, on both tests.

Figure 41.

Average CBAS and paper-and-pencil science achievement and test preference for Icelandic females (top) and males (bottom)

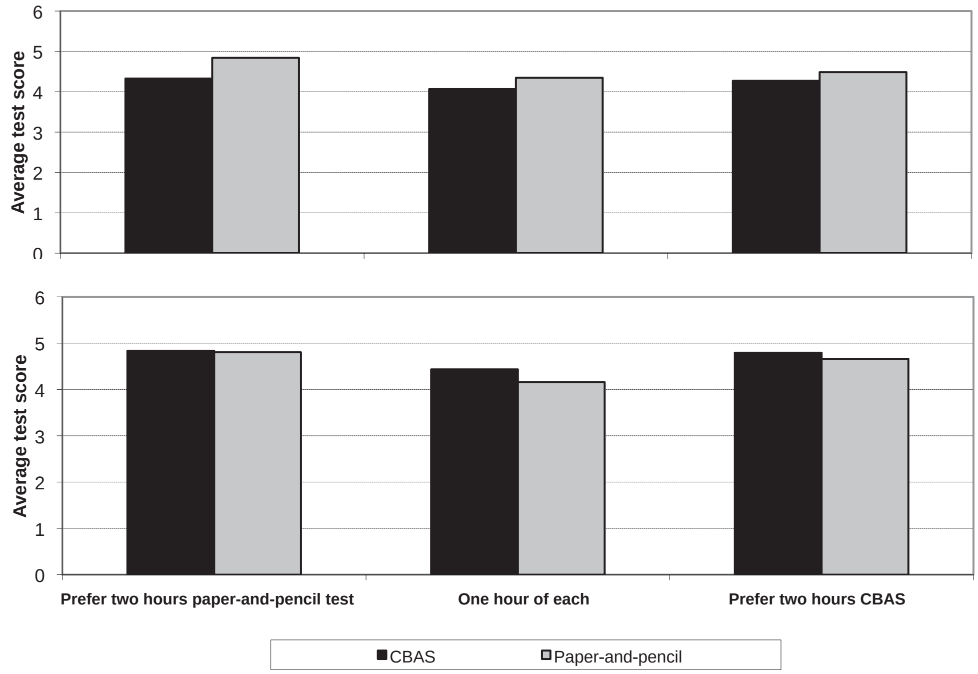


Figure 42.

Average CBAS and paper-and-pencil science achievement and test preference for Korean females (top) and males (bottom)

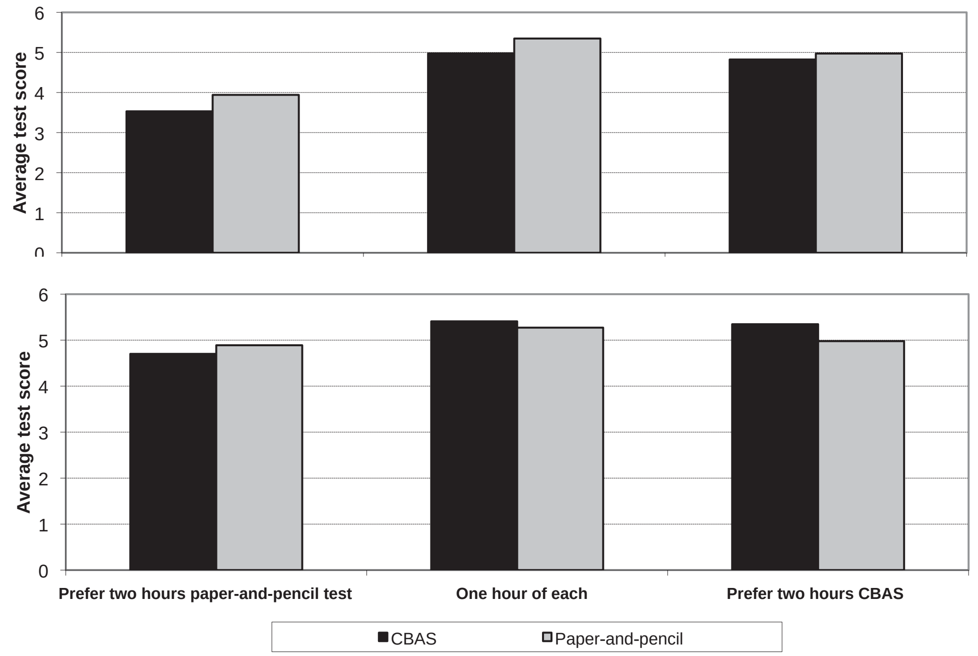


Figure 43.

Average CBAS and paper-and-pencil science achievement and relative effort on tests for Danish females (top) and males (bottom)

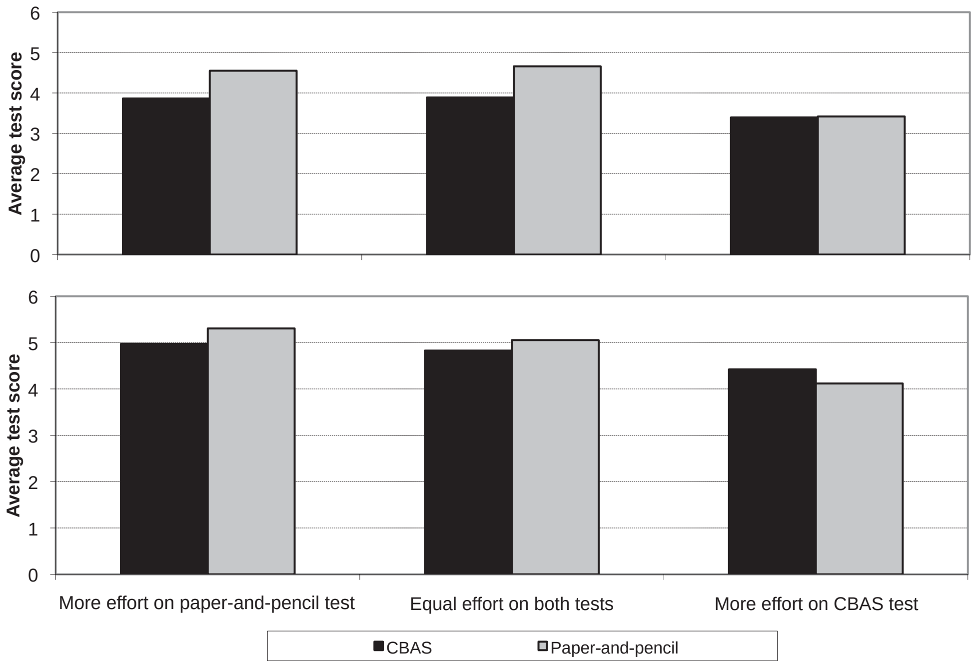


Figure 44.

Average CBAS and paper-and-pencil science achievement and relative effort on tests for Icelandic females (top) and males (bottom)

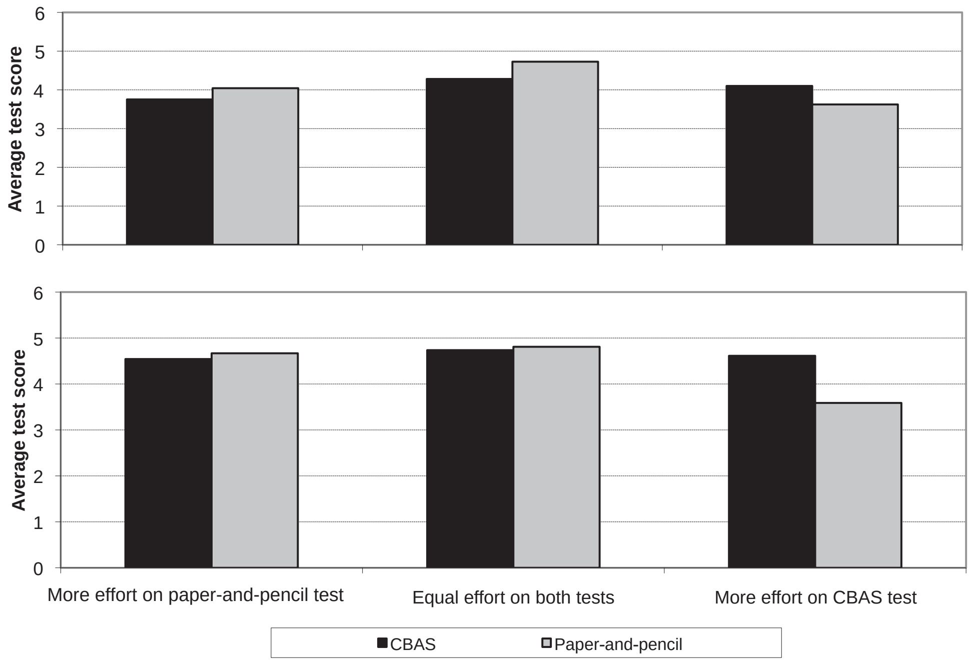
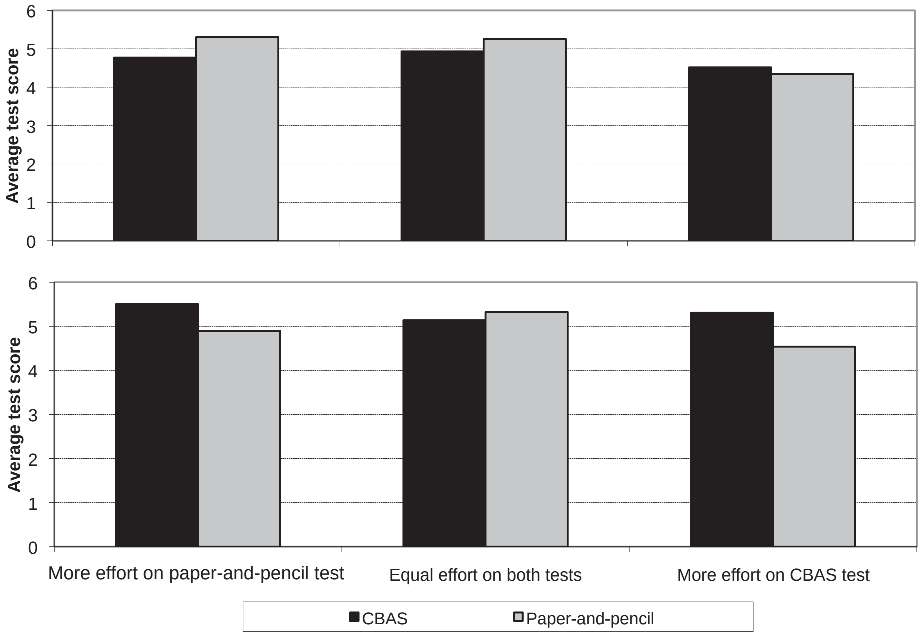


Figure 45.

Average CBAS and paper-and-pencil science achievement and relative effort on tests for Korean females (top) and males (bottom)



As Figure 43 and Figure 44 show, there does not appear to be an association in any country between whether a student worked harder on CBAS or the paper-and-pencil science test and achievement on either test.

STUDENTS' REPORTED EFFORT ON PISA AND CBAS: WHAT THE EFFORT THERMOMETER SHOWS

This section examines reported effort from the PISA effort thermometer across tests, countries and genders, looking at the differences between effort employed in the test and conditional effort reported if the test would have counted for school marks. These findings are considered in relation to the achievement results.

> On average, students reported that they put more effort into the computer-based test than on the paper-and-pencil test.
>
> The difference between actual effort reported and conditional effort reported if the test would have counted for school marks was greater on the paper-and-pencil test than on the computer-based test, particularly for males.
>
> Greater reported effort was associated with higher achievement on the paper-and-pencil test but not on the CBAS test.

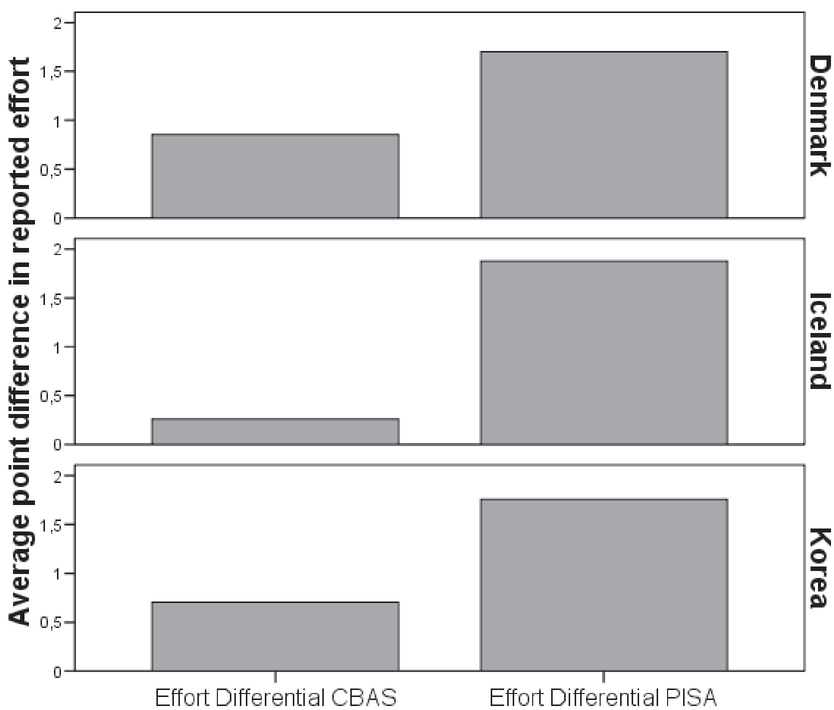
On average, students reported that they tried 5% harder on the CBAS test than on the paper-and-pencil test of science. Considering that the vast majority of student responses were in top 40% of the scale (between 6 and 10 on the thermometer), a 5% increase in effort within this range is considerable. As displayed in Table 22, females in Denmark and Iceland on average reported that they put in more effort than males on both the paper-and-pencil and the computer-based test, whereas the reverse was true in Korea. The amount of effort reported in Iceland overall was inferior to the amounts reported in Denmark and Korea.

Table 22.
Average reported effort per country for the CBAS and the paper-and-pencil tests

Effort thermometer results		Average reported effort on CBAS (out of 10)	S.E. of mean	Average reported effort on PISA (out of 10)	S.E. of mean
Denmark	Females	8.42	0.06	8.18	0.07
	Males	8.01	0.09	7.90	0.08
Iceland	Females	7.78	0.08	7.71	0.09
	Males	7.68	0.10	7.37	0.12
Korea	Females	8.03	0.06	7.41	0.07
	Males	8.12	0.06	7.54	0.07
Total		8.09	0.003	7.54	0.002

The reported effort scores were converted to differential scores which reflect the difference between the real effort and the student's estimate of conditional effort if the test had counted for school marks (as covered in the methodology section). Overall, 73% of students across countries said that they would not have tried any harder if the CBAS test had counted for school marks. This is a very different pattern from the pattern one can observe on the paper-and-pencil effort thermometer scores where only 23% of students said that they would not have put any more effort in if the paper-and-pencil test counted for school marks. A further 23% said they would have tried 1 degree harder on the effort thermometer and an additional 23% said they would have tried 2 degrees harder. As Figure 46 illustrates, on average, PISA effort differential scores were higher

Figure 46.
Effort differential for CBAS and PISA across countries



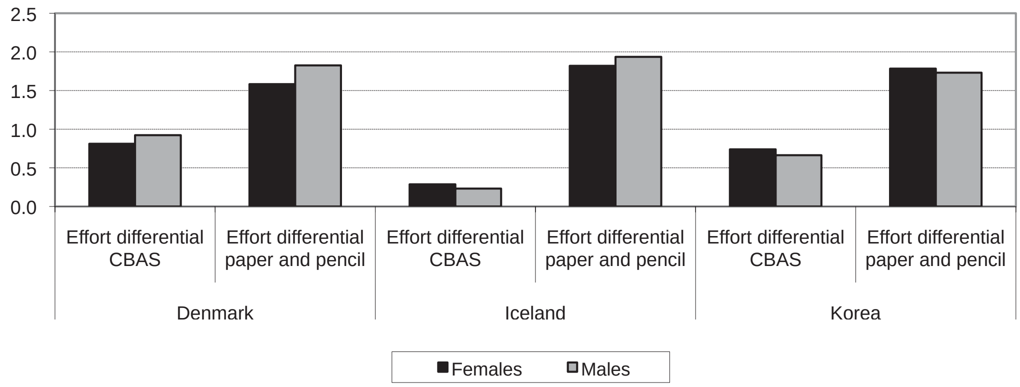
than CBAS differential scores. Given that the CBAS effort scores were higher than the paper-and-pencil effort scores, this indicates either that students were trying close to their maximum on the CBAS test, (perhaps due to greater enjoyment of the test), and thus the scale did not allow them to report a significant increase in effort employed if the test had counted. Another explanation is that due to the novelty of the computer-based testing situation, students found it more difficult to imagine that a computer-based test would really count for final academic marks and as a result their responses do not reflect as accurately the effort they would have manifested. This theory is supported by the higher reported effort in Denmark and Iceland if the CBAS test had counted compared to if the paper-and-pencil test had counted as shown in Table 23 below.

Table 23.
Average reported effort if the CBAS and the paper-and-pencil tests had counted for school marks

	CBAS effort if counted	S.E. of mean	Paper-and-pencil effort if counted	S.E. of mean
Denmark	9.5	0.05	9.7	0.03
Iceland	9.0	0.05	9.3	0.02
Korea	9.4	0.04	9.2	0.03

Figure 47 below shows the gender differences in the gap between reported effort and predicted effort. The gap between reported effort and conditional effort for males was greater than for females in Denmark. In Iceland, the differences are small but show a reversal in that the gap in females' reported effort and conditional effort is greater on CBAS but inferior on paper-and-pencil. In Korea, again the differences are too small to be noteworthy. Overall, only in Denmark it can be confidently said that males report that they would have increased their effort more than females do if the test (either computer-based or paper-and-pencil) had counted.

Figure 47.
How much harder students would have tried if the results of CBAS and the paper-and-pencil test counted for school marks



RELATIONSHIP BETWEEN REPORTED EFFORT AND ACHIEVEMENT

Figure 48 and Figure 49 below display the relationship between effort reported on the effort thermometer and achievement for males and females across countries. Only effort thermometer scores with at least five percent of overall responses are displayed. The figures only a slight tendency towards higher achievement as reported effort increases across all three countries. In contrast, the relationship between paper-and-pencil reported effort and PISA science achievement is clearly shown as a positive relationship; achievement increases with reported effort for both males and females across all three countries.

Figure 48.

CBAS reported effort and achievement across countries

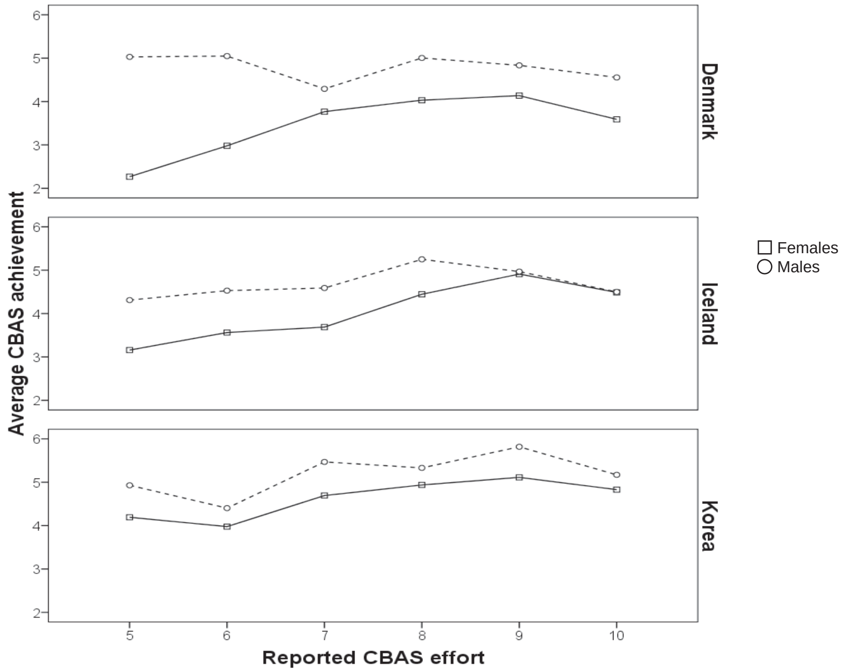
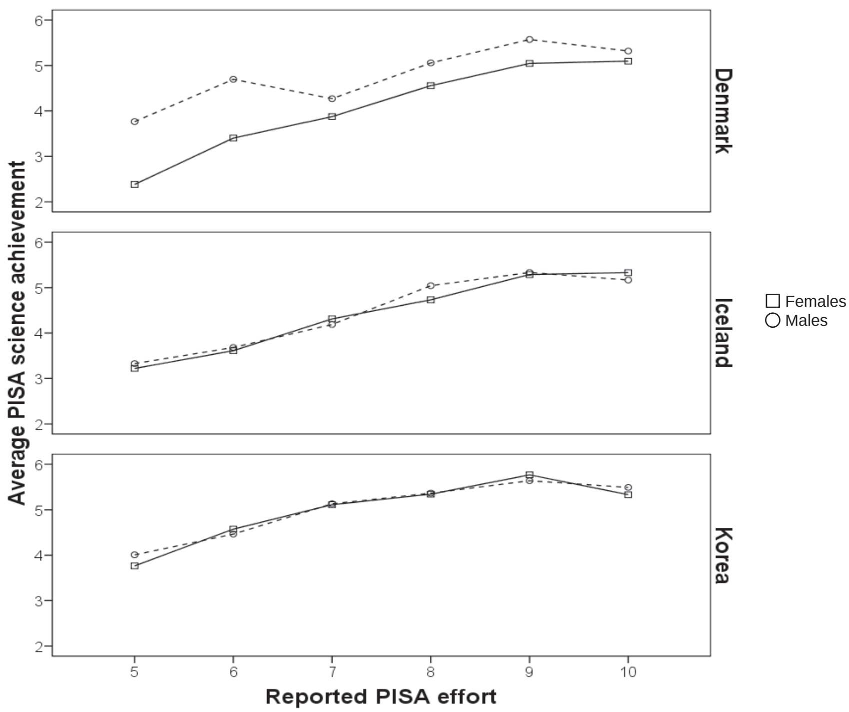


Figure 49.

PISA reported effort and PISA science achievement



The correlation data in Table 24 confirm these trends, showing us that for the paper-and-pencil test, if a student reported that they had put a lot of effort into the test, this was associated with higher performance across all countries and for both males and females. For the CBAS test this relationship was much weaker, particularly for the males.

Table 24.
Correlations between reported effort and achievement

	Paper-and-pencil effort and achievement			CBAS effort and achievement			CBAS reported effort and paper-and-pencil reported effort		
	Females	Males	Total	Females	Males	Total	Females	Males	Total
Denmark	0.40	0.28	0.32	0.14	-0.05	0.01	0.41	0.48	0.46
Iceland	0.42	0.42	0.42	0.31	0.17	0.23	0.37	0.31	0.34
Korea	0.26	0.25	0.25	0.12	0.11	0.12	0.60	0.52	0.55

The results displayed in Table 24 are surprising and warrant further consideration. While higher reported effort predicts better performance on the paper-and-pencil test in all countries across both genders, one finds that the relationship between reported effort and performance on the computer-based test is much weaker, even though the same instrument for measuring effort was utilised (one presented on screen and one presented on paper). Either, students do not respond to the item in the same way when it is presented on screen as compared to on paper, or there is something different in the way that achievement on a computer-based test is related to effort in comparison to the traditional paper-and-pencil tests. The fact that the effort reported on the paper-and-pencil test is a better predictor of CBAS achievement than the effort that students reported on the CBAS[5] test, suggests students respond to the effort thermometer in a different manner (more aligned with the way the item was designed to function) when it is presented on paper than when it is presented on computer. One explanation of this may be that perhaps students are more familiar and thus more accurate reporting their actual effort on a paper-and-pencil test than on a computer test.

As we see in Figure 50 and Figure 51 below, both distributions are negatively skewed with the majority of responses occurring in the top 30% of the effort thermometer. However, since the two distributions are skewed in the same way and approximately to the same degree, this alone cannot explain the large differences in the strength of the relationship between reported effort on the test and achievement on the computer-based or on the paper-and-pencil test.

Figure 50.

Distribution of effort thermometer scores in PISA paper-and-pencil test

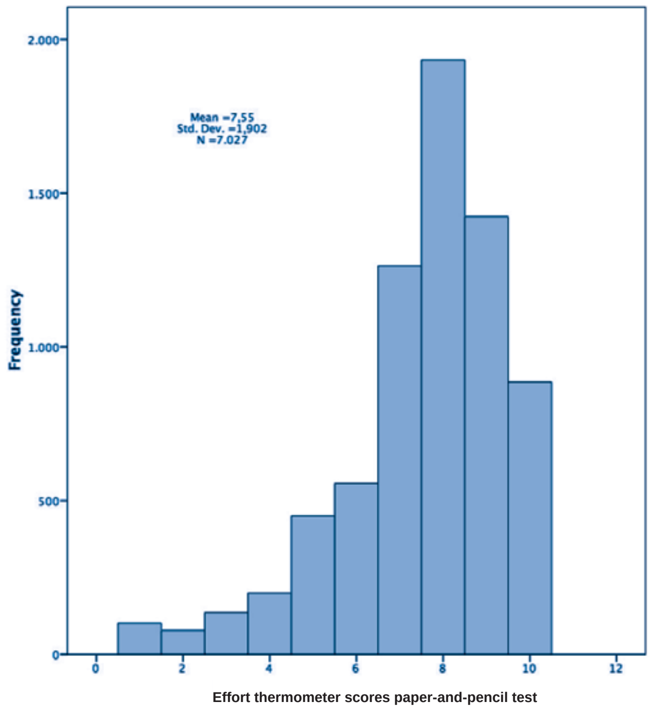
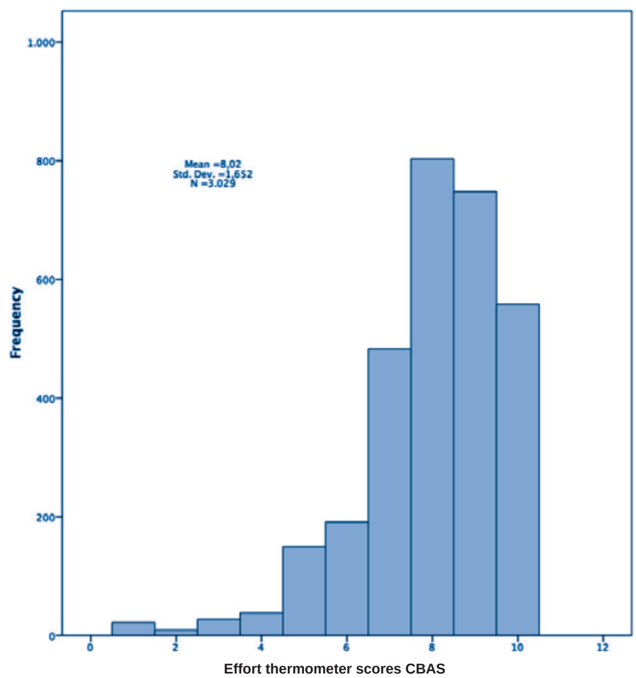


Figure 51.

Distribution of effort thermometer scores in CBAS test



Notes

1. Paper-and-pencil: Denmark (FET =16, p<0.05), Iceland (FET =0.9) , p>0.05) and Korea (FET =12, p<0.05)

2. Denmark (FET =7, p<0.05), Iceland (FET =6, p<0.05) and Korea (FET =1, p>0.05)

3. Denmark (FET =4, p>0.05), Iceland (FET =11, p<0.05) and Korea (FET=23, p<0.05)

4. Denmark (Females: r=0.33, Males: r=0.23), Iceland (Females: r=0.30, Males: r=0.31), Korea (Females: r=0.21, Males: r=0.19)

5. Denmark (Female: r=0.33, Male: r=0.23), Iceland (Female: r=0.30, Male: r=0.31), Korea (Female: r=0.21, Male: r=0.19)

Chapter 6

6

Features of the computer-based items and performance

The high reading load on paper-and-pencil science items appeared as a disadvantage for males. However, males did better than females on more difficult paper-and-pencil items, as well as computer-based items.

While the domain coverage across test modalities was similar, males outperformed females in all domains on the computer-based items. There were more computer-based items that were easier for males than females.

This section looks more closely at specific features of the CBAS items such as the reading load, domain covered and interactivity. To make comparisons of performance across the test modalities, one needs to be confident in assuming that the tests were assessing the same competencies and that differences in performance arise from presentation modality and not from items themselves. Further, one must examine whether there were any gender differences at the item level that may explain overall performance differences.

As covered in Chapter 2, items were presented to students in either one of two forms with item order split at the middle point on the second form. The order scores from the two forms were combined, giving an overall order rank. For example, if an item was 10th on Form 1 and 35th on Form 2, the overall rank score was 45. Analyses were performed to investigate whether the proportion of missing responses and the item difficulty were affected by item presentation order.

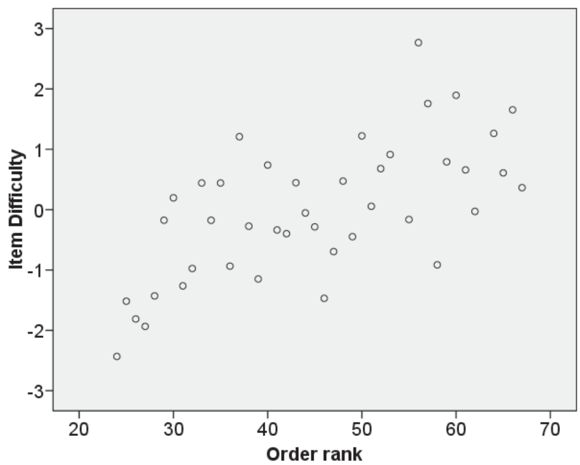
As displayed in Table 25 and Figure 52, item order and item difficulty were found to be positively correlated, indicating that splitting the items over two forms did not entirely adjust for fatigue effects.

Table 25.
Correlations between item order, item difficulty and proportion of missing responses

	Item difficulty	Proportion of missing responses
Item order rank	0.57*	0.71*
Proportion of missing responses	0.48*	--

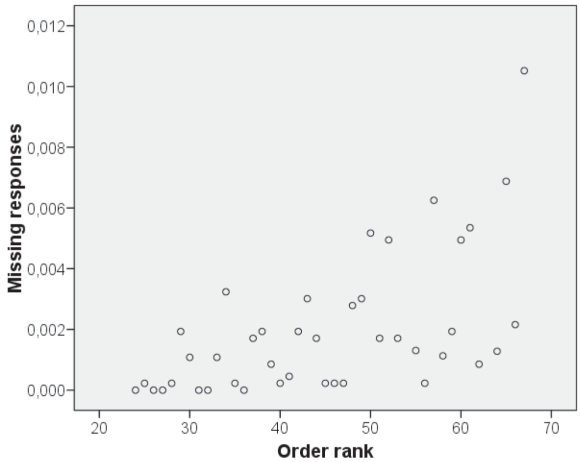
*Significant at p<0.01

Figure 52.
Relationship between item order and item difficulty



Similarly, Figure 53 displays the highly correlated relationship between item order and proportion of missing responses. Again, this is most likely to be caused by fatigue towards the end of the one hour test.

Figure 53.
Relationship between item order and amount of missing responses



MISSING RESPONSES AND GENDER

Percent missing per item was investigated to examine whether males and female display different patterns of missing responses. Before interpreting these results, it should be recalled that overall percentages of missing responses was very low (less than 1%).

On a country-by-country basis, there were a few notable differences. In Denmark, on average females had more missing responses than males (1.08% compared to 0.33%) and on some items, females had over 5% more missing responses. In Iceland the proportions of missing responses overall were low and very similar (0.18% for females compared to 0.16% for males). In Korea missing responses were extremely rare, with males having more than females (0.24% compared to 0.05%). Overall, as Table 26 shows, there were 11 items where females across all countries had 1-2% more missing responses than males.

Table 26.
Items on which females had more missing responses than males

Item	Item Name	Topic
SC819Q01	Farm chemicals	Disease propagation
SC821Q02	Bicycle	Heat from friction
SC853Q03	Tablet Fizz	Water-solid gas reaction
SC951Q01	Plant growth	Experimental trials
SC830Q03	Echolocation	2 videos – perhaps disadvantaged by harder echolocation Q1
SC802Q01	Ping Pong Ball	Expansion of heated air
SC848Q01	Seaside steps	Friction from water
SC829Q01	Chicken and Egg	Chicken birth from egg
SC953Q01	Tower Crane	Forces
SC806Q01	Sliding Continents	Identifying Theory vs. fact
SC830Q01	Echolocation	2 videos – concept of echolocation

Across countries, there were six items where males had on average 1% more missing than females. These are shown in Table 27 below.

Table 27.
Items on which males had more missing responses than females

Item	Item Name	Topic
SC814Q01	Spinifex	Plant reproduction
SC837Q02	Vanishing Krill	Complex item – watch video twice
SC904Q03	Washing glasses	Concept of air expansion in heat – long answers
SC835Q03	Fish Farm	Complex item – photosynthesis
SC837Q01	Vanishing Krill	Complex item – watch video twice and place images in correct boxes
SC843Q01	Travelling with the seasons	Earth's orbit around sun – size of sun

READING LOAD

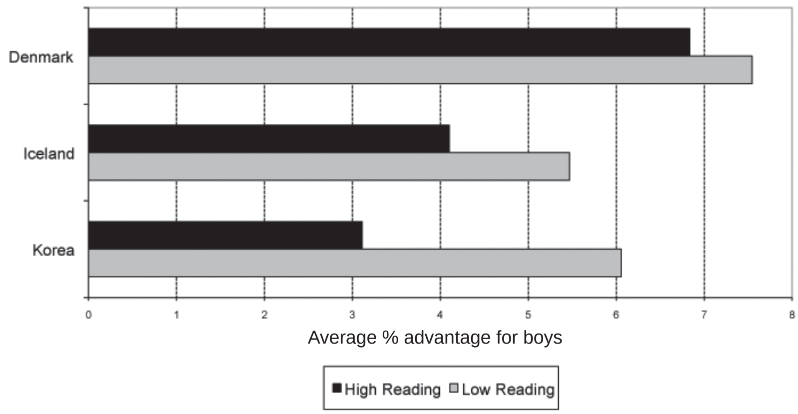
One of the goals of CBAS was to reduce the reading load of the questions, but at the same time retain the science content. It was found that the correlation between the CBAS and PISA reading, at 0.73, was lower than the correlation between PISA science and PISA reading (0.83), so by this measure the goal of reducing the reading load was successful. The following analyses investigate the differences in performance on the items varying in degree of reading load for both CBAS and paper-and-pencil science.

As described in the methodology, all CBAS items were classified as high, medium or low reading load according to the number of words in the item stimulus and question. Percentage correct was calculated for all participating students on the high and low reading load items. Overall, the higher reading load items were more difficult than the lower reading load items, both for males and females across all three countries in both test modalities. One might have expected to see a marked difference in the size of the gender difference between percentages correct on the high and low reading load items. That is, one expected to see a reduction in males' advantage over females when the items were of a higher reading load based on the general PISA trend for females to show higher competency in reading.

Reading load and CBAS achievement across genders

As shown in the following Figure 54, males outperform females on the computer-based items regardless of reading load. This advantage is greater in all three countries for the items of low reading load although the size of the advantage on the low reading load items compared to the high reading load items is relatively small – from under 1% change in Denmark to 3% change in Korea.

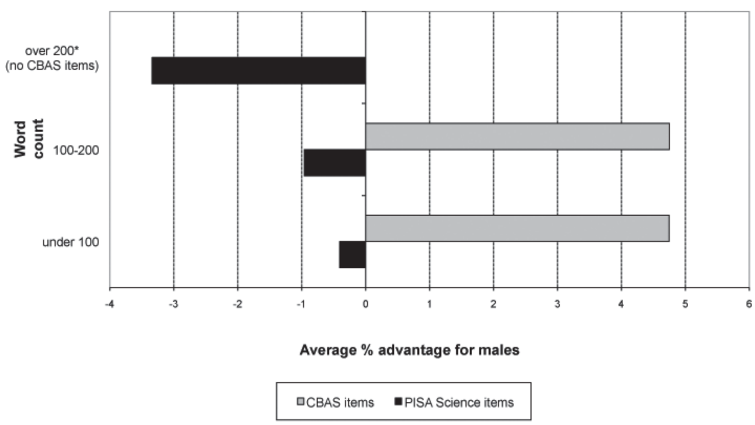
Figure 54.
Average percentage difference in achievement between males and females on high and low reading load CBAS items



Reading load and science achievement for Iceland

In Figure 55, the items have been split into three comparison groups: over 200 words, between 100 and 200 words and under 100 words. As the figure shows, there are no CBAS items that had over 200 words. The results show that, in Iceland, males outperform females on the higher and lower reading load CBAS items. Females outperform the males on all paper-and-pencil items, but by a far greater degree when the items are long. In fact, when the paper-and-pencil items are similar in length to the CBAS items, the gender advantage for females is reduced to less than 1% difference. This is consistent with the results from the PISA 2006 assessment (OECD, 2007a), considering the overall high reading load of the paper-and-pencil items, where Icelandic females outperformed their male counterparts by approximately half a standard deviation. These results indicate that males may be disadvantaged by the length of the paper-and-pencil science items, but they cannot explain fully the advantage for males on the computer-based items.

Figure 55.
Average percentage difference in achievement between males and females according to reading load across test modalities in Iceland



Comparison of very low reading load paper-and-pencil items and CBAS items

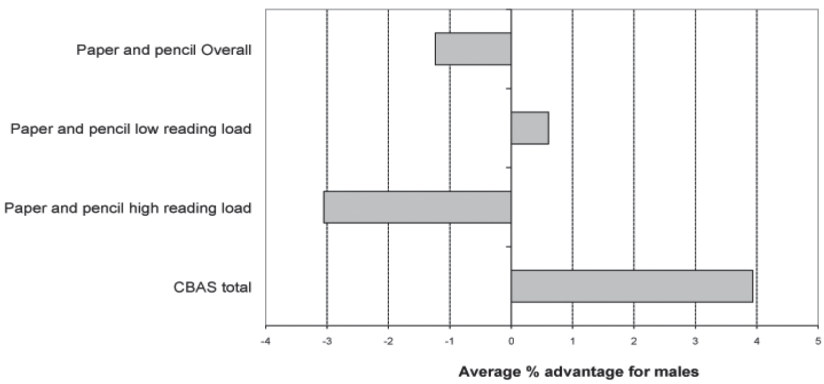
For a more balanced comparison between achievement in CBAS and achievement in paper-and-pencil science according to reading load, only the lowest quarter of items from paper-and-pencil science were selected as the word counts for these items are similar to the word counts for the CBAS items (as shown in Table 28 below).

Table 28.
Item-based word count for CBAS and PISA items

Word count	Paper-and-pencil science[1]	CBAS
Average word count	172	78
Highest	559	140
Average of lowest quarter	87	
25th percentile	117	
Lowest	30	35

Based on the similarities in item length, percentage correct on the paper-and-pencil items with the fewest words (N=24) was compared to percentage correct on all of the CBAS items (N=43) for males and females in Iceland. As we can observe in Figure 56 below, males outperformed females on CBAS items and on the low reading load paper-and-pencil items although the advantage was much greater for the computer-presented items. In contrast, females performed better over all the paper-and-pencil items, an advantage that increased for the particularly high reading load items. These results confirm the proposal that males are significantly disadvantaged by the high reading load of the paper-and-pencil science items as we see a reversal of the advantage for females in the paper-and-pencil science test when only the items with low reading load are considered. However, it is still unclear why the male advantage increases so noticeably when shorter items are presented via the computer.

Figure 56.
Gender difference in percentage correct depending on reading load of items in Iceland



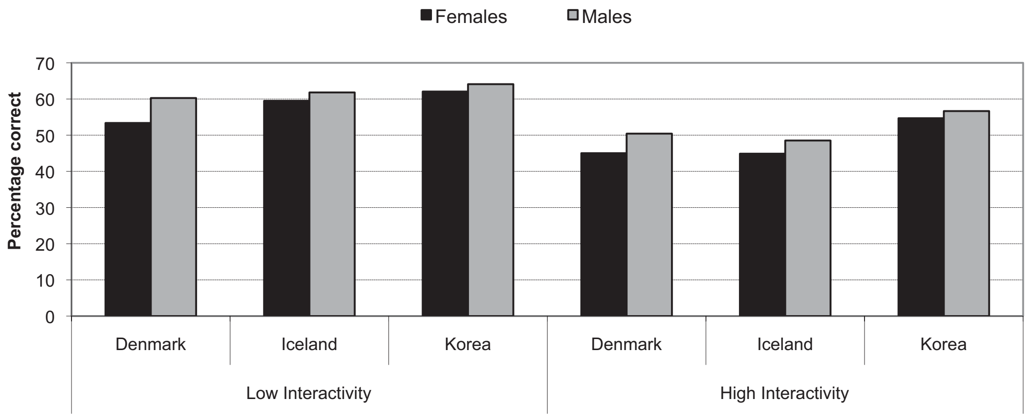
INTERACTIVITY OF ITEMS

One explanation of the gender difference in performance proposed was that males outperformed females on the computer-based items because they were more competent in the types of ICT tasks required of them to complete the items due to their greater ICT familiarity. To investigate this proposal the CBAS items were categorised in terms of their degree of interactivity – for example, whether the item required specific ICT skills such as dragging and dropping or whether it was a relatively simple item involving watching a video and clicking in a response box.

To investigate this, percentage correct was compared for high interactivity items and for low interactivity items. As Figure 57 shows, overall, across both genders and in all three countries, high interactivity items were more difficult than low interactivity items. However, there was no evidence to suggest that the low interactivity items were relatively easier for the females than for the males. This is in contrast to the PISA 2003 ICT report (OECD, 2005b) that showed that the more advanced the ICT tasks became, the wider the gender gap. The absence of a gender gap here indicates that the types of ICT skills necessary to answer these questions are relatively low-level and well within the grasp of most 15-year-old students, regardless of gender.

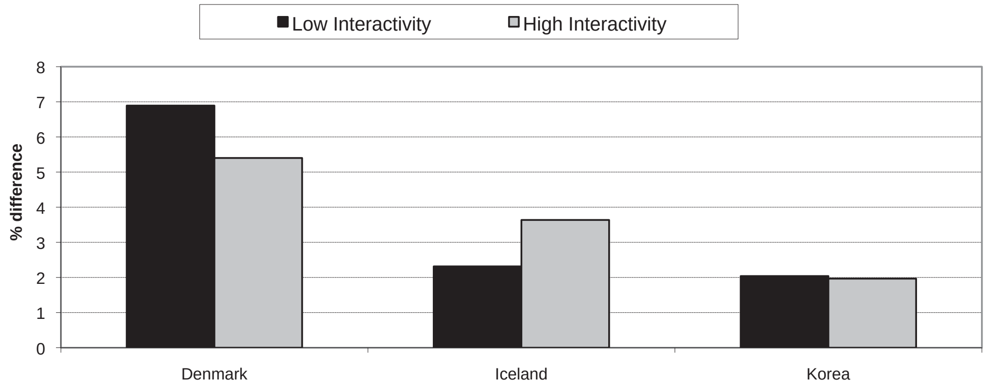
The lower percentage correct overall for high interactivity items may in reality be an artefact of the item type in that the more complex items, *e.g.* items involving dragging dropping or trialling experiments, required more complex, often two part, answers, calling upon higher reasoning which students had a higher chance of getting wrong.

Figure 57.
Percentage correct on high and low interactivity CBAS items



Presented in another way in Figure 58 below, the gender difference in performance does not clearly increase or decrease according to the interactivity of the items. Males clearly outperform females on high and low interactivity items. However, it should be noted that the 'interactivity' of the CBAS items was relatively low overall as these items were designed to be easy for even the most ICT unfamiliar students to successfully complete. A study such as the PISA 2009 Electronic Reading Assessment with highly interactive items simulating an on-line searching environment will provide researchers with a greater range of item interactivity to examine any potential impacts on performance.

Figure 58.
Average performance advantage for males across countries according to the interactivity of CBAS items.



GENDER DIFFERENCES IN PERCENTAGE CORRECT PER ITEM

As mentioned in the methodology section, several items were identified as easier for females or males in the gender differential item functioning analysis after the Field Trial in 2005. Using the percentage correct results, we examined whether these specific items behaved in the same way in the Main Study and whether there were any additional items displaying gender bias. The purpose here was to identify what sorts of items males and females did better on; if reading load, item interactivity and item difficulty could not completely explain the differences in performance, perhaps closer consideration of the item content could account for some of the differences.

Of the three items identified as easier for males in the DIF analyses, all three showed a clear bias towards males when measured by percentage correct. Males performed 18% and 11% better than females on the two bicycle-related questions overall and 22% better than females on the item about the eagle (Designed for Flight). In addition, the second question in the Ping Pong Ball unit was approximately 10% easier for males in all three countries, the item about a car assembly line was easier for males in Denmark and Iceland (18% and 12% higher correct responses respectively) but 5% easier for females in Korea and the item regarding growing grain was 10% easier for males overall.

Of the two items that were identified as easier for females in the DIF analyses, the same biases were observed with percentage correct, but the size of the difference in percentage correct was much smaller than for the items where the males outperformed the females. Ping Pong Ball Q1 was 3% easier for females in Denmark and 5% easier for females in Iceland and Washing Dishes Q3 was 3% easier in all countries. In addition, the Chicken and Egg Q2 was 5% easier for females overall and females outperformed males on the Nuclear Power Plant item by 7%.

It is interesting to note here that while males outperformed females on Q2 of the Ping Pong Ball unit, females slightly outperformed males (in Denmark and Iceland) on the first question of the same unit while the general them of the items are common. These items are displayed in Figure 59 and Figure 60 below.

Figure 59.

Sample unit: Ping Pong Ball Q1

Question 37: Ping Pong Ball

Chris places his dented ping pong ball in hot water and holds it under the water.

Which one of the following statements best explains why the dent in the ball disappears?

- ○ The ball material expands and stretches over the dent.
- ○ The ball material shrinks and pushes the dent inside out.
- ○ More molecules are created inside the ball.
- ○ Molecules inside the ball move faster and push the dent out.

Figure 60.

Sample unit: Ping Pong Ball Q2

Question 38: Ping Pong Ball

The movie shows a ping pong ball placed in the air stream of a hair dryer.

Why doesn't the ping pong ball go higher in the air stream?

- ○ Its weight equals the push of the air.
- ○ Air resistance pushes it down.
- ○ It is stuck in a vacuum.
- ○ The air stream stops at that point.

The following Figure 61 and Figure 62 display two items on which males perform better than females. The first question relates to riding a bicycle (and in particular purposefully skidding to a stop) which is perhaps an activity that males have more experience with than females. The second item relates to an eagle preparing to land which involves themes of aerodynamics and birds of prey that may be more familiar to males than females.

Figure 61.
Sample item showing bias towards male: Bicycle Q1

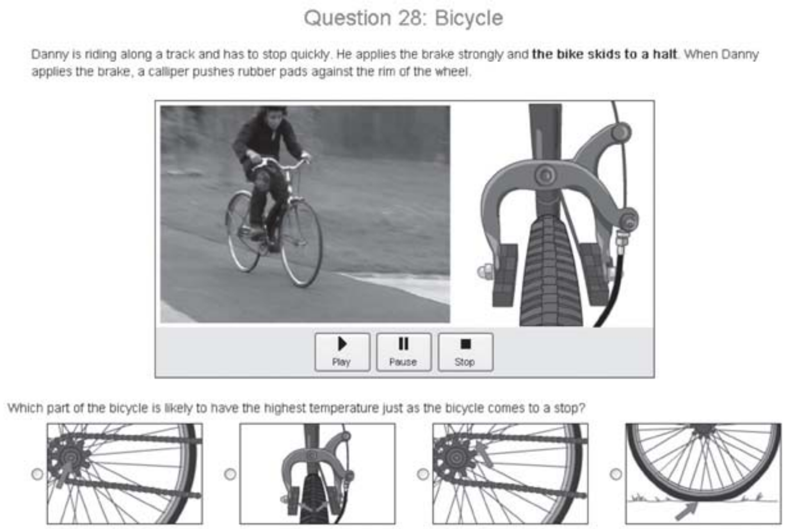
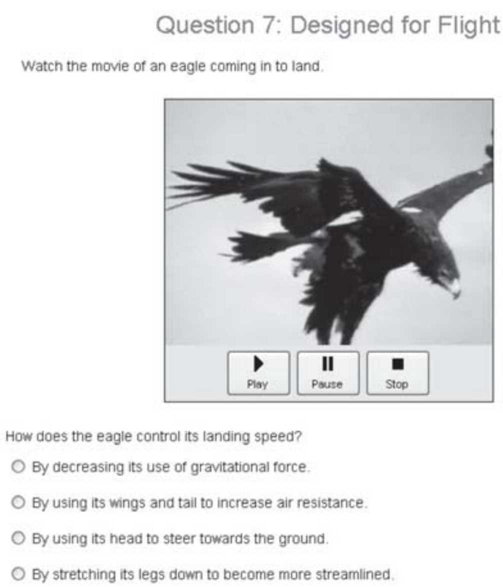


Figure 62.
Sample item showing bias towards male: Designed for Flight Q1



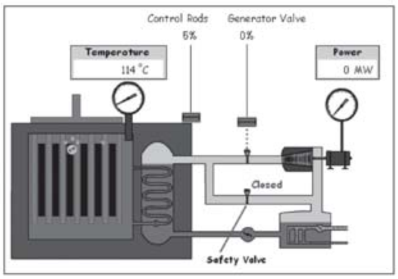
The following Figure 63 and Figure 64 display two items on which females outperform males. The first question relates to the Nuclear Power Plant and was a very interactive item requiring the student to monitor power production based on changes they made to the control rods and generator valve settings and the second item relates to an experiment to do with an amount of bubbles produced by different amounts of detergent. It is difficult to characterise these items in a specific way that would partially explain why females found these items easier than males did.

Figure 63.

Sample item showing bias towards female: Nuclear Power Plant

Question 36: Nuclear Power Plant

In the nuclear power plant simulation below, the number of nuclear interactions produced by the reactor is regulated by the **Control Rods**. The power output is regulated by the **Generator Valve**. The **Control Rods** and the **Generator Valve** can be controlled by dragging the sliders up and down.



Set the generator valve at 80%. How far (%) should the control rods be lifted for the power plant to supply a continuous output of 800 Megawatts (MW) without the safety valve opening?

- 55%
- 70%
- 85%
- 100%

Figure 64.

Sample item showing bias towards female: Washing Glasses Q3

Question 20: Washing Glasses

John was curious about the formation of bubbles when he washed glasses. He experimented with three identical drinking glasses washed in three different ways.

What question was John trying to answer by conducting his experiment with the three drinking glasses?

- Does the position on the draining board change the number of bubbles formed?
- Does the temperature of the glass affect the number of bubbles being formed?
- Do bubbles forming inside a glass differ from those forming outside a glass?
- Do different amounts of detergent produce different sized bubbles?

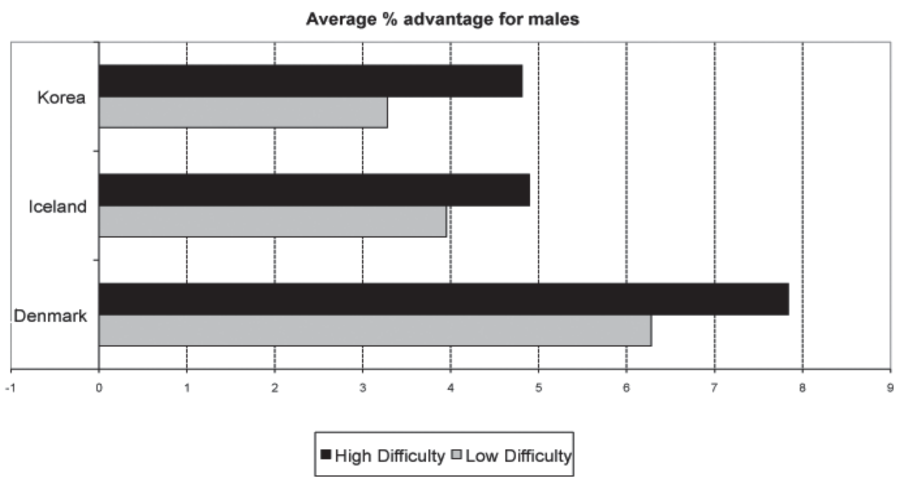
On the whole, the items which showed an advantage for males cannot easily be classified as easier due to low reading load, nor due to higher interactivity. Further, it does not appear to be the domains assessed that affect whether females or males will do better on the item, nor the medium of presentation (animation, video footage, still image, etc). The only potential explanations on an item-by-item basis appear to be the themes assessed.

An overall explanation of why the males outperformed the females may be offered by the finding that overall the computer-based items were easier than the paper-based items. Percent correct is much higher overall for CBAS than for paper-and-pencil science for all countries. It is possible that the difficulty of the paper-and-pencil science items fatigues the males and encourages them to 'give up' more so than females. The following analyses investigate whether easier CBAS items were comparatively easier for males or females.

Gender differences in performance and item difficulty

Using the PISA item parameters for CBAS items as outlined in the methodology, all CBAS items were classified into three groups according to their item difficulty score: high, medium and low. Percentage correct for males and for females was calculated for the low and high groups and the average difference between the males' percentage correct and the females' percentage was calculated and is displayed in Figure 65 below.

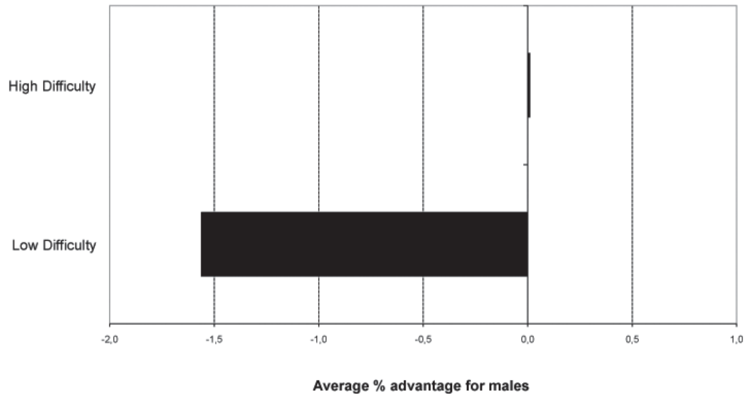
Figure 65.
Percentage correct advantage for males on high and low difficulty CBAS items.



As we see in Figure 65, there is a clear advantage in percentage correct for males in all three countries regardless of the difficulty of the CBAS item. Yet, we do see evidence to support the theory that the males do comparatively better on the easier items than females. In other words, males' performance advantage is greater for the high difficulty items (they get 5.8% more correct on these items than females) than on the low difficulty items (where they get on average 4.5% more correct than females).

While this pattern is notable, it is not, however, unique to the computer-based assessment and we can observe a similar pattern in the Icelandic paper-and-pencil science achievement results. Icelandic males do comparatively better on the more difficult paper-based science items. Whereas females outperform males on the low difficulty items (and overall), there are no performance differences on the high difficulty items. These patterns are shown in Figure 66.

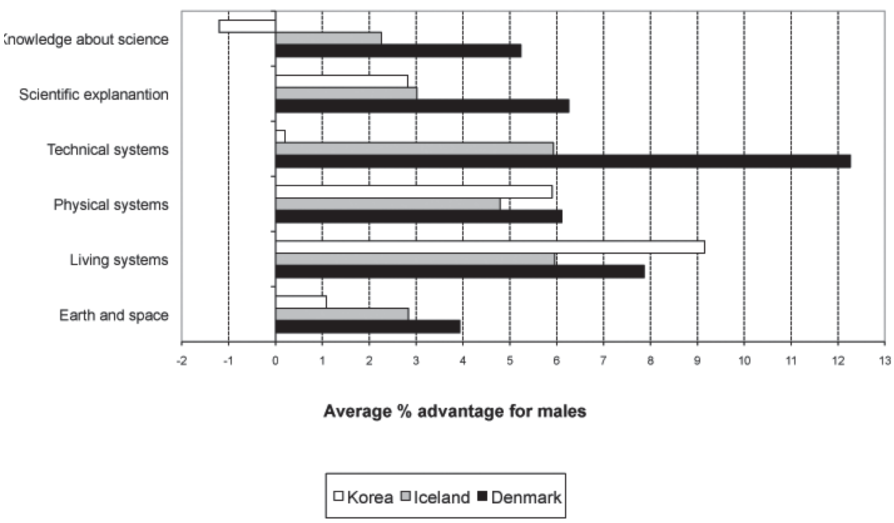
Figure 66.
Gender difference in performance between harder and easier paper-and-pencil items



ACHIEVEMENT ACROSS SUB DOMAINS OF SCIENCE

In order to further examine the similarities and differences between the paper-and-pencil and the computer tests, the following section of this report explores males' and females' performance per domain of science across the tests. According to the PISA 2006 Science Framework (OECD, 2006), physical systems items should favour males. Yet initial CBAS Field Trial analyses (PISA Project Consortium, 2005) and the following Figure 67 indicate that the physics character of the items does not appear related to the gender difference in performance. In fact, males outperform females in the CBAS items in all domains in Denmark and Iceland and in all but one domain in Korea (Scientific Enquiry – Knowledge about science).

Figure 67.
Gender difference in performance across domains of science in CBAS

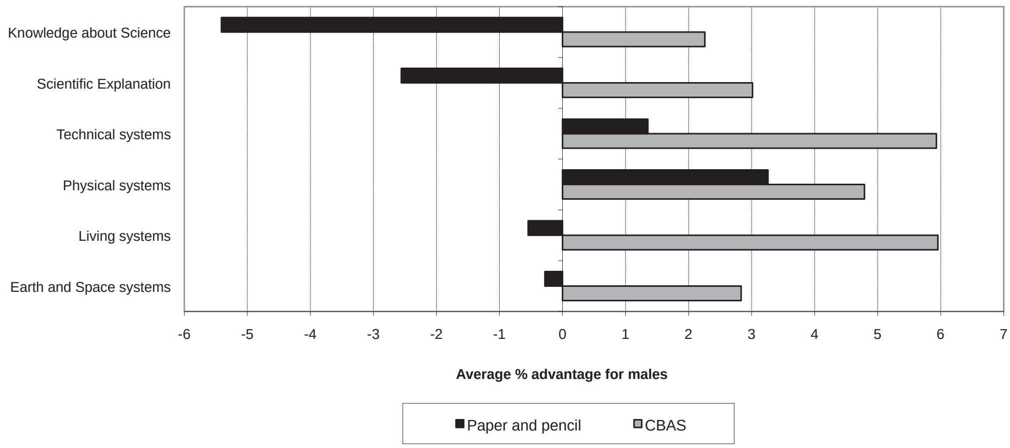


Paper-and-pencil science performance across domains in Iceland

In the paper-and-pencil PISA 2006 results (OECD, 2007a) it was reported that females performed significantly better overall than males on the Knowledge about Science items (which combine both the Scientific Explanation and the Scientific Enquiry items). This general pattern was also present in the Icelandic data shown in Figure 68 below using a percent correct calculation. Here we see that females outperform males on the items assessing the methods of science (Knowledge about Science), whereas overall males have the advantage on the Knowledge of Science items (despite slight advantages for females in the Living systems and Earth and Space systems questions).

Figure 68.

Average advantage in performance across domains for CBAS and paper-and-pencil science items in Iceland



When we compare the performance in domains across test modalities, it is interesting to note that the male advantage decreases in the same domains that the females displayed strengths in on the paper-and-pencil test (Knowledge about Science). This is a good indication that the items across both tests are assessing the same competencies as the gender difference changes in the same manner across domains regardless of test modality.

Notes

1. Note that the PISA word count per item is slightly over estimated as the unit stimulus need only be read once but is included in the word count per item. However, for the following analyses this should not affect the patterns analysed.

Chapter 7

7

Summary

Paper-and-pencil 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.*

Gender differences: *On the paper-and-pencil test of science in Iceland, females outperformed the males and in Denmark males outperformed the females. However, when the test was administered via computer, males in all three countries outperformed their female counterparts.*

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

Test motivation and enjoyment: *Students on the whole enjoyed the computer-based test more than the paper-and-pencil test 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 the case for the computer-based test.*

By far the most clear cut and most interesting finding from these analyses is the finding that, whereas overall country-by-country performance levels did not change between tests, males in all three countries outperformed the females when the test was presented via computer. This gap between the males' performance and the females' performance occurred regardless of the patterns of achievement across gender on the paper-and-pencil test of science (recall that in the paper-and-pencil test males outperformed females in Denmark, females outperformed males in Iceland and there were no gender differences in Korea).

EXPLAINING THE GENDER DIFFERENCES

The increase in males' performance may at least partially be explained by the lower reading load as when performance on the paper-and-pencil items that were similar in length to the CBAS items was compared with performance on CBAS, we noted a substantial increase in performance from the males and the gender advantage for females is completely removed.

The questionnaire results may also shed some light on explaining why the males perform better on CBAS. In particular:

- Males are more motivated on the CBAS test and they enjoy it more,
- Males have more experience with computer-based games, Internet, games-type software that would be similar to the flash animations and video footage used for the CBAS items,
- Males are more confident on ICT tasks and may therefore approach the test with more confidence,
- Males use computers outside the home more than females which may contribute to greater confidence in skills transference and greater familiarity with different keyboards, screens and software.

However, despite the intuitive relationship between higher motivation, greater experience with and confidence for ICT tasks and achievement on the computer-based test, statistical analysis of the correlations between achievement and all of these questionnaire factors did not reveal any significant associations between ICT use factors and achievement. Consequently, we must consider other factors that may have influenced performance.

Was this "a test for males"?

Upon closer investigation of the types of items presented in the computer test, it appears that there may be a bias in the gender-typing of the items with a strong content bias towards males in the video footages used. For example, there are nine videos over five units showing males performing certain activities, (riding bikes, throwing litter in the bin, etc), where the males are specifically named in the text and sighted in the video footage. There are a further two items in one unit where a male is named and illustrated as the principal actor in the animated scene. On the females' side, there are no items showing females performing activities by video and only one item that refers to a female by name. The lack of females in the items may lead to a lower level of engagement with the test for the females and a consequently lower level of performance.

Three notes of caution are necessary before concluding this section. Firstly, while the overall achievement results were very clear-cut and the gender difference in favour of males very obvious in each country, finding explanations for the achievement results in the responses to the questionnaire was more difficult due to high levels of variations between countries. The small number of countries involved in this study should be kept in mind when interpreting these results and in order to further clarify patterns of changes in performance when testing is presented via computer, further cross-national research will be necessary.

Secondly, a comment must be made about the nature of the experimental design employed here which compares achievement on the paper-based test of science to achievement on the computer-based test of science using a completely different set of items, involving different subjects and item types. Because both of these tests were developed within the same PISA 2006 framework and the final set of items approved by similar groups of subject area experts, it has been necessary to assume the equivalence of the tests to for analyses in several sections of this report. However, in some instances this assumption may not be valid, in that the computer-based test: a) was one hour shorter than the paper-and-pencil test, b) involved a completely different set of items assessing knowledge about different types of scientific topics and c) involved different item types than the paper-and-pencil test, for example, scientific trial and error items where students were able to make certain choices and trial their consequences before responding to the item. As a result of these differences, it is very difficult to conclude whether the performance differences observed are a consequence of the change in presentation modality, the different items, the different item types, or a combination of all factors.

For a more balanced test design, valuable insight would be provided in the future by conducting a similar experiment using a third reference group where a group of matched students are given the same paper-and-pencil items via computer. This would enable researchers to separate out more clearly the effects of changing test modality, changing test format and changing the test items themselves.

Finally, the mention of rapidly changing practices regarding ICT use in education introduced at the beginning of the report must be reiterated here. At the time at which this study was performed, for most students across all three countries this was the first time they had ever done a 'test' on computer and there was considerable novelty to the situation. In contrast, the controlled environment test-taking conditions for the paper-and-pencil test were similar to regular test-taking conditions in schools which are often associated with anxiety and stress for students. The differences between the relationship between effort reported and achievement on the paper-and-pencil test and the (lack of) relationship between effort report and achievement on the computer-based test point towards something inherently different about how students approached the computer-based test and how they approached the paper-and-pencil test. We must ask whether they were able to 'imagine that the results counted for overall school marks' as the instrument requires them to do as easily on the computer-based test as they could on the paper-and-pencil test. Does the fact that they find the test more enjoyable mean that they also perceive it as more trivial? The effects of the novelty or, perhaps, lack of gravity of the computer-based test in students' perceptions should be monitored over time as it may fade rapidly as students become more accustomed to sitting tests with implications for the future via computer. For example in Denmark, national testing via computer has already begun so students will already be familiar with receiving important academic results from a computer-based test.

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. In general, changing the test modality to a computer-based presentation platform should not affect performance at the country level, however the current results indicate that it will negatively impact the performance of females in comparison to the males. When presenting tests via computer, students may report higher levels of enjoyment, effort and may prefer the computer-based test to a paper-based test but this preference does not mean that achievement will be higher. These domains should be investigated further by national testing institutes wishing to adapt their testing systems, and in particular for Iceland, the reversal of the pattern of achievement by gender and the strange relationship between ICT familiarity and achievement for the Icelandic females requires detailed future enquiry.

References

Bridgeman, B., Lennon, M.L., and **Jackenthal, A.** (2001), *Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Performance,* Educational Testing Service, Research Report No. 23.

Gallagher, A., Bridgeman, B., and **Cahalan, C.** (2000), *The effect of computer-based tests on racial/ethnic, gender, and language groups,* Educational Testing Service, Research Report No. 00-8, US.

Harrison, C., Comber, C., Fisher, T., Haw, K., Lewin, C., Lunzer, E., McFarlane, A., Mavers, Di., Scrimshaw, P., Somekh, B., and **Watling, R.** (2004), *ImpaCT2: The impact of information and communications technology on pupil learning and attainment,* Department for Education and Skills, UK.

Hoffman, C. (2008, June 26), "The Battle For Facebook", Rolling Stone, [http://www.rollingstone.com/news/story/21129674/the_battle_for_facebook - Electronic version accessed 08/07/2008].

Mason, B.J., *et al*. (2001), An examination of the equivalence between non-adaptive computer-based and traditional testing, *Journal of Educational Computing Research*, No. 24, Vol. 1,pp. 29-40.

OECD (2004), *Learning for Tomorrow's World – First Results from PISA 2003*, OECD Publishing.

OECD (2005a), *PISA 2003 Data Analysis Manual: SPSS Users,* OECD Publishing.

OECD (2005b), *Are Students Ready for a Technology-Rich World? What PISA Studies Tell Us*, OECD Publishing.

OECD (2006), *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006,* OECD Publishing.

OECD (2007a), *Science Competencies for Tomorrow's World. Volume 1: Analysis*, OECD Publishing.

OECD (2007b), *Science Competencies for Tomorrow's World. Volume 2: Data*, OECD Publishing.

OECD (2009), *PISA 2006 Technical Report*, OECD Publishing.

PISA Project Consortium (2005), "CBAS Preliminary Field Trial Data Analysis," Meeting paper presented at National Project Maneger Meeting, Mildura, Australia.

Ravitz, J., Mergendoller, J. and **Rush, W.** (2002), "Cautionary tales about correlations between student computer use and academic achievement", Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Singleton, C. (2001), Computer-based assessment in education, *Educational & Child Psychology*, No. 18, Vol 3, pp. 58-74.

Taylor, C., Jamieson, J., Eignor, D., and **Kirsch, I**. (1998), "The relationship between computer familiarity and performance on computer-based TOEFL tasks", *TOEFL research report 59*, Educational Testing Service.

Valentine, G., Marsh, J., Pattie, C. and **BMRB** (2005), "Children and young people's home use of ICT for educational purposes: The impact on attainment", Department for Education and Skills Research Report RR672, University of Leeds, UK.

Wilhelm, O. and **Schroeders, U.** (2008), Computerized ability measurement: Some substantive Dos and Don'ts*, in Towards a Research Agenda on Computer-Based Assessment*, European Commission Joint Research Centre, Italy.

Zandvliet, D. and **Farragher, P.** (1997), "A comparison of the computer and written forms of multiple-choice testing" *Journal of Research on Computing in Education*, No. 29, pp. 423-438.

Annex A

SAMPLING INFORMATION

The decision to impute plausible values on the CBAS scale for students in CBAS participating schools that participated only in PISA paper-and-pencil test was made by data analysts of the Consortium (Westat and ACER). The quality of imputing missing values was evaluated and the ISL CBAS situation approximated as closely as possible: effective sample size = 400, correlation between SCIE and CBAS r=0.9, correlation between probability to respond and CBAS performance p = 0.1 (it is low, because for most students the non-response is random, so for that group p = 0) and non-response rate = 0.8. In the following simulation, only scores on one other dimension are used (therefore background information, reading and mathematics performance are not taken into account). The results of this imputation are in the following table:

ARGUMENTS	CBAS mean of respondents only	S.E. resp	True mean CBAS	S.E. true	Mean CBAS after imputation	S.E. impute	DIFF (true-impute)	S.E. DIFF
N=400, r=0.9, p=0.1, NR=0.8	**0.08**	0.106	**0.01**	0.049	**0.00**	0.073	0.01	0.054

As the above table shows, the mean after imputation is much closer to the true mean (and not significantly different) than the mean of the 20% respondents, even with such a low p-value. The bias of the 20% of the students will be bigger is we increase p to 0.2, while the mean after imputation is still the same as the true mean:

ARGUMENTS	CBAS mean of respondents only	S.E. resp	True mean CBAS	S.E. true	Mean CBAS after imputation	S.E. impute	DIFF (true-impute)	S.E. DIFF
N=400, r=0.9, p=0.2, NR=0.8	**0.28**	0.103	**0.01**	0.049	**0.01**	0.073	0.00	0.055

Here, the CBAS mean of the 20% respondents is significantly different from 0, while the true mean is not and neither is the mean after imputation.

Annex B

FREQUENCY OF PERFORMING ICT ACTIVITIES BY COUNTRY AND GENDER

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Used computer IC1		4.8%	5.8%	5.3%
		1 Yes	95.2%	93.2%	94.2%
		2 No		1.0%	0.5%
	Total		100.0%	100.0%	100.0%
Iceland	Used computer IC1		1.3%	3.2%	2.3%
		1 Yes	98.6%	95.5%	97.0%
		2 No	0.1%	1.3%	0.7%
	Total		100.0%	100.0%	100.0%
Korea	Used computer IC1		0.4%	0.3%	0.3%
		1 Yes	99.5%	99.4%	99.4%
		2 No	0.2%	0.3%	0.3%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How long used computers IC2		4.9%	5.9%	5.4%
		1 Less than 1 year	0.6%	0.8%	0.7%
		2 1 to 3 years	6.3%	4.7%	5.5%
		3 3 to 5 years	24.3%	14.3%	19.4%
		4 5 years or more	63.9%	74.2%	69.0%
	Total		100.0%	100.0%	100.0%
Iceland	How long used computers IC2		1.3%	3.3%	2.3%
		1 Less than 1 year	0.6%	0.6%	0.6%
		2 1 to 3 years	7.9%	8.3%	8.1%
		3 3 to 5 years	28.3%	18.9%	23.6%
		4 5 years or more	62.0%	68.9%	65.4%
	Total		100.0%	100.0%	100.0%
Korea	How long used computers IC2		0.4%	0.3%	0.3%
		1 Less than 1 year	1.3%	0.8%	1.0%
		2 1 to 3 years	4.4%	3.3%	3.8%
		3 3 to 5 years	13.9%	13.1%	13.5%
		4 5 years or more	80.0%	82.6%	81.4%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Use computer at home IC3a		4.9%	5.9%	5.4%
		1 Almost every day	76.7%	87.9%	82.2%
		2 Once or twice a week	11.8%	5.4%	8.6%
		3 Few times a month	4.0%	0.4%	2.2%
		4 Once a month or less	1.4%	0.1%	0.8%
		5 Never	1.2%	0.3%	0.7%
	Total		100.0%	100.0%	100.0%
Iceland	Use computer at home IC3a		1.3%	3.3%	2.3%
		1 Almost every day	85.5%	90.9%	88.2%
		2 Once or twice a week	9.3%	4.2%	6.8%
		3 Few times a month	2.0%	0.6%	1.3%
		4 Once a month or less	1.1%	0.5%	0.8%
		5 Never	0.8%	0.6%	0.7%
	Total		100.0%	100.0%	100.0%
Korea	Use computer at home IC3a		0.4%	0.3%	0.3%
		1 Almost every day	55.6%	65.6%	61.2%
		2 Once or twice a week	35.8%	28.2%	31.6%
		3 Few times a month	6.0%	3.4%	4.6%
		4 Once a month or less	0.6%	0.4%	0.5%
		5 Never	1.7%	2.1%	1.9%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Use computer at school IC3b		4.9%	5.9%	5.4%
		1 Almost every day	14.9%	21.0%	17.9%
		2 Once or twice a week	43.6%	44.9%	44.3%
		3 Few times a month	30.4%	21.0%	25.8%
		4 Once a month or less	5.0%	6.1%	5.6%
		5 Never	1.1%	1.0%	1.1%
	Total		100.0%	100.0%	100.0%
Iceland	Use computer at school IC3b		1.3%	3.3%	2.3%
		1 Almost every day	5.2%	9.1%	7.1%
		2 Once or twice a week	41.4%	47.6%	44.5%
		3 Few times a month	23.7%	19.3%	21.5%
		4 Once a month or less	17.5%	11.7%	14.6%
		5 Never	11.0%	9.1%	10.0%
	Total		100.0%	100.0%	100.0%
Korea	Use computer at school IC3b		0.4%	0.3%	0.3%
		1 Almost every day	5.6%	2.8%	4.0%
		2 Once or twice a week	34.6%	34.1%	34.4%
		3 Few times a month	10.4%	9.9%	10.1%
		4 Once a month or less	12.5%	10.7%	11.5%
		5 Never	36.6%	42.2%	39.7%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Use computer other places IC3c		4.9%	6.0%	5.5%
		1 Almost every day	7.2%	8.7%	8.0%
		2 Once or twice a week	16.5%	25.3%	20.9%
		3 Few times a month	26.5%	29.7%	28.1%
		4 Once a month or less	23.6%	16.9%	20.3%
		5 Never	21.2%	13.4%	17.3%
	Total		100.0%	100.0%	100.0%
Iceland	Use computer other places IC3c		1.3%	3.3%	2.3%
		1 Almost every day	5.4%	9.3%	7.3%
		2 Once or twice a week	28.1%	31.8%	29.9%
		3 Few times a month	29.5%	27.3%	28.5%
		4 Once a month or less	22.5%	17.3%	19.9%
		5 Never	13.1%	11.0%	12.1%
	Total		100.0%	100.0%	100.0%
Korea	Use computer other places IC3c		0.4%	0.3%	0.3%
		1 Almost every day	2.8%	6.3%	4.7%
		2 Once or twice a week	12.7%	35.0%	25.1%
		3 Few times a month	20.2%	31.4%	26.4%
		4 Once a month or less	25.1%	14.9%	19.4%
		5 Never	38.8%	12.1%	24.0%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Browse Internet IC4a		4.9%	5.9%	5.4%
		1 Almost every day	14.5%	22.5%	18.5%
		2 Once or twice a week	40.8%	38.6%	39.7%
		3 Few times a month	28.0%	19.9%	24.0%
		4 Once a month or less	7.7%	7.8%	7.8%
		5 Never	4.1%	5.3%	4.7%
	Total		100.0%	100.0%	100.0%
Iceland	Browse Internet IC4a		1.3%	3.2%	2.3%
		1 Almost every day	38.7%	45.8%	42.3%
		2 Once or twice a week	34.8%	29.9%	32.3%
		3 Few times a month	16.4%	12.6%	14.5%
		4 Once a month or less	5.6%	4.5%	5.1%
		5 Never	3.2%	4.0%	3.6%
	Total		100.0%	100.0%	100.0%
Korea	Browse Internet IC4a		0.4%	0.3%	0.3%
		1 Almost every day	23.5%	22.5%	22.9%
		2 Once or twice a week	44.0%	43.3%	43.6%
		3 Few times a month	21.1%	20.1%	20.5%
		4 Once a month or less	7.8%	9.4%	8.7%
		5 Never	3.3%	4.5%	4.0%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Play games IC4b		4.9%	6.0%	5.4%
		1 Almost every day	9.2%	52.6%	30.7%
		2 Once or twice a week	23.1%	25.2%	24.1%
		3 Few times a month	18.2%	9.3%	13.8%
		4 Once a month or less	17.4%	4.3%	10.9%
		5 Never	27.1%	2.6%	15.0%
	Total		100.0%	100.0%	100.0%
Iceland	Play games IC4b		1.3%	3.3%	2.3%
		1 Almost every day	10.7%	48.6%	29.6%
		2 Once or twice a week	22.6%	26.6%	24.5%
		3 Few times a month	22.8%	11.2%	17.0%
		4 Once a month or less	21.6%	5.9%	13.8%
		5 Never	21.1%	4.5%	12.8%
	Total		100.0%	100.0%	100.0%
Korea	Play games IC4b		0.4%	0.3%	0.3%
		1 Almost every day	9.2%	38.6%	25.5%
		2 Once or twice a week	22.5%	37.7%	30.9%
		3 Few times a month	14.6%	14.9%	14.8%
		4 Once a month or less	20.0%	4.2%	11.3%
		5 Never	33.3%	4.3%	17.3%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Write documents IC4c		4.9%	5.9%	5.4%
		1 Almost every day	12.7%	13.4%	13.0%
		2 Once or twice a week	45.6%	44.9%	45.2%
		3 Few times a month	30.5%	29.0%	29.7%
		4 Once a month or less	3.2%	5.4%	4.3%
		5 Never	3.0%	1.5%	2.3%
	Total		100.0%	100.0%	100.0%
Iceland	Write documents IC4c		1.3%	3.3%	2.3%
		1 Almost every day	6.5%	7.9%	7.2%
		2 Once or twice a week	27.0%	27.5%	27.3%
		3 Few times a month	45.3%	39.3%	42.3%
		4 Once a month or less	13.9%	15.6%	14.7%
		5 Never	6.1%	6.5%	6.3%
	Total		100.0%	100.0%	100.0%
Korea	Write documents IC4c		0.4%	0.3%	0.3%
		1 Almost every day	4.9%	4.3%	4.6%
		2 Once or twice a week	28.1%	24.0%	25.8%
		3 Few times a month	36.4%	37.2%	36.8%
		4 Once a month or less	22.2%	23.7%	23.0%
		5 Never	8.0%	10.5%	9.4%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Collaborate on Internet IC4d		4.9%	6.0%	5.4%
		1 Almost every day	6.3%	11.0%	8.6%
		2 Once or twice a week	22.4%	25.5%	24.0%
		3 Few times a month	25.8%	26.8%	26.3%
		4 Once a month or less	17.9%	10.9%	14.4%
		5 Never	22.6%	19.8%	21.2%
	Total		100.0%	100.0%	100.0%
Iceland	Collaborate on Internet IC4d		1.3%	3.3%	2.3%
		1 Almost every day	6.1%	13.5%	9.8%
		2 Once or twice a week	12.8%	17.6%	15.2%
		3 Few times a month	28.6%	26.5%	27.6%
		4 Once a month or less	27.2%	20.4%	23.8%
		5 Never	24.0%	18.7%	21.4%
	Total		100.0%	100.0%	100.0%
Korea	Collaborate on Internet IC4d		0.4%	0.3%	0.3%
		1 Almost every day	7.7%	7.9%	7.8%
		2 Once or twice a week	15.8%	19.3%	17.7%
		3 Few times a month	23.4%	21.5%	22.3%
		4 Once a month or less	25.1%	25.7%	25.4%
		5 Never	27.7%	25.3%	26.4%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Use spreadsheets IC4e		5.0%	5.9%	5.5%
		1 Almost every day	1.7%	5.9%	3.8%
		2 Once or twice a week	10.3%	17.4%	13.8%
		3 Few times a month	21.1%	27.9%	24.5%
		4 Once a month or less	28.1%	20.0%	24.1%
		5 Never	33.7%	22.9%	28.4%
	Total		100.0%	100.0%	100.0%
Iceland	Use spreadsheets IC4e		1.3%	3.3%	2.3%
		1 Almost every day	1.8%	4.3%	3.0%
		2 Once or twice a week	5.1%	9.2%	7.1%
		3 Few times a month	18.3%	23.2%	20.7%
		4 Once a month or less	30.4%	27.4%	28.9%
		5 Never	43.1%	32.7%	37.9%
	Total		100.0%	100.0%	100.0%
Korea	Use spreadsheets IC4e		0.4%	0.3%	0.3%
		1 Almost every day	1.4%	2.6%	2.1%
		2 Once or twice a week	5.9%	7.6%	6.8%
		3 Few times a month	13.7%	13.9%	13.8%
		4 Once a month or less	26.2%	27.7%	27.0%
		5 Never	52.5%	47.9%	49.9%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Download software IC4f		5.0%	6.0%	5.4%
		1 Almost every day	5.7%	25.0%	15.3%
		2 Once or twice a week	12.8%	28.5%	20.6%
		3 Few times a month	14.8%	16.3%	15.5%
		4 Once a month or less	16.1%	9.8%	13.0%
		5 Never	45.6%	14.4%	30.1%
	Total		100.0%	100.0%	100.0%
Iceland	Download software IC4f		1.3%	3.3%	2.3%
		1 Almost every day	7.9%	31.5%	19.6%
		2 Once or twice a week	13.6%	26.1%	19.8%
		3 Few times a month	20.7%	18.8%	19.7%
		4 Once a month or less	23.1%	9.8%	16.5%
		5 Never	33.3%	10.6%	22.0%
	Total		100.0%	100.0%	100.0%
Korea	Download software IC4f		0.4%	0.3%	0.3%
		1 Almost every day	7.4%	17.8%	13.2%
		2 Once or twice a week	20.9%	34.9%	28.6%
		3 Few times a month	24.3%	25.1%	24.7%
		4 Once a month or less	22.1%	14.7%	18.0%
		5 Never	24.9%	7.3%	15.2%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Graphics programs		4.9%	5.9%	5.4%
	IC4g	1 Almost every day	5.3%	10.4%	7.9%
		2 Once or twice a week	12.5%	15.0%	13.8%
		3 Few times a month	19.3%	21.5%	20.4%
		4 Once a month or less	20.8%	17.4%	19.1%
		5 Never	37.1%	29.7%	33.4%
	Total		100.0%	100.0%	100.0%
Iceland	Graphics programs		1.3%	3.3%	2.3%
	IC4g	1 Almost every day	6.5%	11.3%	8.9%
		2 Once or twice a week	14.9%	13.7%	14.3%
		3 Few times a month	20.5%	22.6%	21.5%
		4 Once a month or less	26.5%	20.4%	23.5%
		5 Never	30.3%	28.7%	29.5%
	Total		100.0%	100.0%	100.0%
Korea	Graphics programs		0.4%	0.3%	0.3%
	IC4g	1 Almost every day	6.6%	4.6%	5.5%
		2 Once or twice a week	15.6%	10.1%	12.5%
		3 Few times a month	18.4%	16.5%	17.4%
		4 Once a month or less	24.2%	27.3%	25.9%
		5 Never	34.9%	41.2%	38.4%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Educational software IC4h		5.0%	5.9%	5.4%
		1 Almost every day	1.4%	2.9%	2.2%
		2 Once or twice a week	7.6%	10.5%	9.1%
		3 Few times a month	10.0%	17.3%	13.7%
		4 Once a month or less	21.7%	22.3%	22.0%
		5 Never	54.2%	41.0%	47.7%
	Total		100.0%	100.0%	100.0%
Iceland	Educational software IC4h		1.3%	3.3%	2.3%
		1 Almost every day	2.0%	5.4%	3.7%
		2 Once or twice a week	5.8%	9.2%	7.5%
		3 Few times a month	16.7%	20.3%	18.5%
		4 Once a month or less	26.3%	23.3%	24.8%
		5 Never	48.0%	38.5%	43.2%
	Total		100.0%	100.0%	100.0%
Korea	Educational software IC4h		0.4%	0.3%	0.3%
		1 Almost every day	3.2%	4.0%	3.6%
		2 Once or twice a week	10.6%	10.0%	10.3%
		3 Few times a month	14.9%	12.8%	13.7%
		4 Once a month or less	19.5%	23.4%	21.6%
		5 Never	51.4%	49.6%	50.4%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Download music IC4i		4.9%	6.0%	5.5%
		1 Almost every day	20.7%	35.0%	27.7%
		2 Once or twice a week	21.4%	20.9%	21.2%
		3 Few times a month	13.8%	10.2%	12.0%
		4 Once a month or less	9.7%	7.1%	8.4%
		5 Never	29.6%	20.9%	25.3%
	Total		100.0%	100.0%	100.0%
Iceland	Download music IC4i		1.3%	3.3%	2.3%
		1 Almost every day	29.3%	46.4%	37.8%
		2 Once or twice a week	27.0%	24.5%	25.8%
		3 Few times a month	17.2%	8.4%	12.8%
		4 Once a month or less	9.1%	6.2%	7.6%
		5 Never	16.2%	11.3%	13.7%
	Total		100.0%	100.0%	100.0%
Korea	Download music IC4i		0.4%	0.3%	0.3%
		1 Almost every day	32.9%	33.3%	33.1%
		2 Once or twice a week	39.5%	42.9%	41.4%
		3 Few times a month	18.7%	16.5%	17.5%
		4 Once a month or less	5.3%	4.0%	4.6%
		5 Never	3.3%	3.0%	3.1%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	Write programs IC4j		4.9%	5.9%	5.4%
		1 Almost every day	2.3%	7.1%	4.7%
		2 Once or twice a week	2.6%	7.6%	5.0%
		3 Few times a month	4.6%	9.8%	7.2%
		4 Once a month or less	9.2%	12.2%	10.7%
		5 Never	76.4%	57.4%	67.0%
	Total		100.0%	100.0%	100.0%
Iceland	Write programs IC4j		1.3%	3.3%	2.3%
		1 Almost every day	1.7%	9.3%	5.5%
		2 Once or twice a week	4.1%	8.3%	6.2%
		3 Few times a month	10.6%	15.3%	12.9%
		4 Once a month or less	19.4%	17.3%	18.4%
		5 Never	62.9%	46.6%	54.8%
	Total		100.0%	100.0%	100.0%
Korea	Write programs IC4j		0.4%	0.3%	0.3%
		1 Almost every day	1.7%	3.5%	2.7%
		2 Once or twice a week	4.1%	6.2%	5.3%
		3 Few times a month	7.2%	10.8%	9.2%
		4 Once a month or less	16.1%	19.7%	18.1%
		5 Never	70.5%	59.6%	64.5%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	E-mail or chat rooms		4.9%	6.0%	5.4%
	IC4k	1 Almost every day	53.7%	51.1%	52.4%
		2 Once or twice a week	20.8%	20.9%	20.8%
		3 Few times a month	7.9%	8.0%	8.0%
		4 Once a month or less	3.9%	5.7%	4.8%
		5 Never	8.7%	8.3%	8.5%
	Total		100.0%	100.0%	100.0%
Iceland	E-mail or chat rooms		1.3%	3.3%	2.3%
	IC4k	1 Almost every day	74.0%	69.6%	71.8%
		2 Once or twice a week	16.6%	15.9%	16.3%
		3 Few times a month	4.3%	4.2%	4.2%
		4 Once a month or less	1.7%	3.0%	2.4%
		5 Never	2.2%	4.0%	3.1%
	Total		100.0%	100.0%	100.0%
Korea	E-mail or chat rooms		0.4%	0.3%	0.3%
	IC4k	1 Almost every day	34.9%	38.9%	37.1%
		2 Once or twice a week	28.7%	31.6%	30.3%
		3 Few times a month	14.4%	12.9%	13.5%
		4 Once a month or less	10.7%	5.9%	8.0%
		5 Never	10.9%	10.5%	10.7%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Chat IC5a		4.9%	5.9%	5.4%
		1 Do well by myself	85.4%	85.1%	85.3%
		2 Do with help	6.3%	4.3%	5.3%
		3 Know but can't do	3.0%	3.3%	3.2%
		4 Don't know	0.4%	1.3%	0.8%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Chat IC5a		1.3%	3.3%	2.3%
		1 Do well by myself	96.1%	92.2%	94.1%
		2 Do with help	1.2%	2.2%	1.7%
		3 Know but can't do	1.0%	1.3%	1.1%
		4 Don't know	0.4%	1.1%	0.7%
	Total		100.0%	100.0%	100.0%
Korea	How well - Chat IC5a		0.4%	0.3%	0.3%
		1 Do well by myself	92.7%	93.1%	93.0%
		2 Do with help	3.1%	2.9%	3.0%
		3 Know but can't do	3.1%	2.6%	2.8%
		4 Don't know	0.7%	1.1%	0.9%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Virus IC5b		5.0%	5.9%	5.4%
		1 Do well by myself	17.1%	62.6%	39.6%
		2 Do with help	35.4%	21.2%	28.4%
		3 Know but can't do	38.2%	8.7%	23.6%
		4 Don't know	4.4%	1.6%	3.0%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Virus IC5b		1.3%	3.3%	2.3%
		1 Do well by myself	27.3%	68.2%	47.6%
		2 Do with help	30.8%	16.6%	23.7%
		3 Know but can't do	35.0%	9.8%	22.5%
		4 Don't know	5.6%	2.1%	3.9%
	Total		100.0%	100.0%	100.0%
Korea	How well - Virus IC5b		0.4%	0.3%	0.3%
		1 Do well by myself	43.0%	62.3%	53.7%
		2 Do with help	36.8%	24.4%	29.9%
		3 Know but can't do	14.1%	9.1%	11.4%
		4 Don't know	5.7%	3.8%	4.7%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Edit photos IC5c		5.0%	6.0%	5.5%
		1 Do well by myself	52.2%	59.2%	55.7%
		2 Do with help	30.5%	26.4%	28.5%
		3 Know but can't do	11.0%	7.0%	9.0%
		4 Don't know	1.4%	1.3%	1.4%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Edit photos IC5c		1.3%	3.3%	2.3%
		1 Do well by myself	63.7%	63.2%	63.5%
		2 Do with help	23.6%	21.2%	22.4%
		3 Know but can't do	9.7%	9.7%	9.7%
		4 Don't know	1.7%	2.6%	2.2%
	Total		100.0%	100.0%	100.0%
Korea	How well - Edit photos IC5c		0.4%	0.3%	0.3%
		1 Do well by myself	62.9%	51.4%	56.6%
		2 Do with help	26.6%	33.0%	30.1%
		3 Know but can't do	8.3%	12.2%	10.4%
		4 Don't know	1.9%	3.2%	2.6%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Database IC5d		5.0%	5.9%	5.5%
		1 Do well by myself	6.9%	16.3%	11.6%
		2 Do with help	22.6%	33.4%	28.0%
		3 Know but can't do	34.4%	27.7%	31.1%
		4 Don't know	31.1%	16.6%	23.9%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Database IC5d		1.3%	3.3%	2.3%
		1 Do well by myself	11.4%	24.5%	17.9%
		2 Do with help	22.2%	27.1%	24.6%
		3 Know but can't do	33.0%	27.3%	30.2%
		4 Don't know	32.1%	17.8%	25.0%
	Total		100.0%	100.0%	100.0%
Korea	How well - Database IC5d		0.4%	0.3%	0.3%
		1 Do well by myself	9.0%	11.1%	10.2%
		2 Do with help	34.4%	38.1%	36.4%
		3 Know but can't do	22.9%	26.3%	24.8%
		4 Don't know	33.2%	24.3%	28.3%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Copy data to CD IC5e		5.0%	5.9%	5.4%
		1 Do well by myself	63.5%	81.6%	72.5%
		2 Do with help	19.2%	6.8%	13.0%
		3 Know but can't do	9.9%	3.9%	6.9%
		4 Don't know	2.5%	1.7%	2.1%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Copy data to CD IC5e		1.3%	3.3%	2.3%
		1 Do well by myself	69.4%	82.1%	75.8%
		2 Do with help	17.9%	9.1%	13.5%
		3 Know but can't do	9.6%	3.8%	6.7%
		4 Don't know	1.8%	1.7%	1.8%
	Total		100.0%	100.0%	100.0%
Korea	How well - Copy data to CD IC5e		0.4%	0.3%	0.3%
		1 Do well by myself	34.2%	54.8%	45.6%
		2 Do with help	33.3%	26.0%	29.3%
		3 Know but can't do	24.5%	13.5%	18.4%
		4 Don't know	7.7%	5.5%	6.5%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Move files IC5f		5.0%	5.9%	5.4%
		1 Do well by myself	75.6%	88.1%	81.8%
		2 Do with help	14.4%	4.1%	9.3%
		3 Know but can't do	4.5%	1.4%	3.0%
		4 Don't know	0.5%	0.4%	0.5%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Move files IC5f		1.3%	3.3%	2.3%
		1 Do well by myself	75.3%	86.4%	80.8%
		2 Do with help	13.0%	6.0%	9.5%
		3 Know but can't do	8.2%	2.9%	5.5%
		4 Don't know	2.3%	1.4%	1.9%
	Total		100.0%	100.0%	100.0%
Korea	How well - Move files IC5f		0.4%	0.3%	0.3%
		1 Do well by myself	85.4%	90.0%	88.0%
		2 Do with help	9.3%	6.7%	7.9%
		3 Know but can't do	2.5%	1.2%	1.8%
		4 Don't know	2.4%	1.8%	2.0%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Search Internet IC5g		5.0%	5.9%	5.4%
		1 Do well by myself	92.6%	88.2%	90.5%
		2 Do with help	1.8%	4.5%	3.1%
		3 Know but can't do	0.5%	0.5%	0.5%
		4 Don't know	0.2%	0.8%	0.5%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Search Internet IC5g		1.3%	3.3%	2.3%
		1 Do well by myself	95.6%	91.0%	93.3%
		2 Do with help	2.2%	3.8%	3.0%
		3 Know but can't do	0.7%	1.0%	0.9%
		4 Don't know	0.2%	0.9%	0.6%
	Total		100.0%	100.0%	100.0%
Korea	How well - Search Internet IC5g		0.4%	0.3%	0.3%
		1 Do well by myself	98.2%	96.4%	97.2%
		2 Do with help	0.7%	2.4%	1.6%
		3 Know but can't do	0.3%	0.6%	0.5%
		4 Don't know	0.5%	0.3%	0.4%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Download files IC5h		5.0%	5.9%	5.4%
		1 Do well by myself	58.0%	81.9%	69.8%
		2 Do with help	22.8%	9.7%	16.3%
		3 Know but can't do	12.1%	1.6%	6.9%
		4 Don't know	2.1%	0.9%	1.5%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Download files IC5h		1.3%	3.3%	2.3%
		1 Do well by myself	55.7%	82.4%	69.0%
		2 Do with help	25.3%	8.3%	16.8%
		3 Know but can't do	14.4%	4.3%	9.4%
		4 Don't know	3.3%	1.7%	2.5%
	Total		100.0%	100.0%	100.0%
Korea	How well - Download files IC5h		0.4%	0.3%	0.3%
		1 Do well by myself	95.8%	95.6%	95.7%
		2 Do with help	2.7%	2.6%	2.7%
		3 Know but can't do	0.6%	0.5%	0.5%
		4 Don't know	0.5%	1.0%	0.8%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Attach e-mail IC5i		5.0%	5.9%	5.4%
		1 Do well by myself	80.2%	78.9%	79.5%
		2 Do with help	9.1%	9.4%	9.3%
		3 Know but can't do	4.3%	4.0%	4.1%
		4 Don't know	1.5%	1.8%	1.6%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Attach e-mail IC5i		1.3%	3.3%	2.3%
		1 Do well by myself	74.3%	78.5%	76.4%
		2 Do with help	15.4%	11.5%	13.5%
		3 Know but can't do	6.5%	4.3%	5.4%
		4 Don't know	2.5%	2.5%	2.5%
	Total		100.0%	100.0%	100.0%
Korea	How well - Attach e-mail IC5i		0.4%	0.3%	0.3%
		1 Do well by myself	96.9%	94.3%	95.5%
		2 Do with help	1.7%	3.4%	2.6%
		3 Know but can't do	0.5%	0.7%	0.6%
		4 Don't know	0.6%	1.3%	1.0%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Word processor IC5j		5.0%	5.9%	5.4%
		1 Do well by myself	87.6%	83.9%	85.7%
		2 Do with help	6.0%	7.1%	6.5%
		3 Know but can't do	0.9%	2.7%	1.8%
		4 Don't know	0.6%	0.4%	0.5%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Word processor IC5j		1.3%	3.3%	2.3%
		1 Do well by myself	90.9%	84.9%	87.9%
		2 Do with help	4.8%	6.4%	5.6%
		3 Know but can't do	1.9%	3.9%	2.9%
		4 Don't know	1.1%	1.5%	1.3%
	Total		100.0%	100.0%	100.0%
Korea	How well - Word processor IC5j		0.4%	0.3%	0.3%
		1 Do well by myself	92.0%	85.8%	88.6%
		2 Do with help	6.1%	9.8%	8.1%
		3 Know but can't do	0.8%	2.2%	1.5%
		4 Don't know	0.8%	2.0%	1.5%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Spreadsheet IC5k		5.0%	6.0%	5.5%
		1 Do well by myself	44.9%	58.9%	51.9%
		2 Do with help	35.1%	24.3%	29.7%
		3 Know but can't do	13.0%	8.8%	10.9%
		4 Don't know	2.0%	2.0%	2.0%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Spreadsheet IC5k		1.3%	3.3%	2.3%
		1 Do well by myself	32.2%	49.3%	40.7%
		2 Do with help	33.3%	28.5%	30.9%
		3 Know but can't do	24.2%	13.5%	18.9%
		4 Don't know	9.0%	5.4%	7.2%
	Total		100.0%	100.0%	100.0%
Korea	How well - Spreadsheet IC5k		0.4%	0.3%	0.3%
		1 Do well by myself	20.2%	21.3%	20.8%
		2 Do with help	42.0%	42.6%	42.3%
		3 Know but can't do	19.6%	20.7%	20.2%
		4 Don't know	17.8%	15.1%	16.3%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Presentation IC5l		5.0%	5.9%	5.4%
		1 Do well by myself	61.2%	66.7%	63.9%
		2 Do with help	22.3%	17.8%	20.1%
		3 Know but can't do	9.1%	6.8%	7.9%
		4 Don't know	2.4%	2.8%	2.6%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Presentation IC5l		1.3%	3.3%	2.3%
		1 Do well by myself	62.1%	64.2%	63.1%
		2 Do with help	20.6%	18.8%	19.7%
		3 Know but can't do	10.6%	9.9%	10.2%
		4 Don't know	5.5%	3.8%	4.6%
	Total		100.0%	100.0%	100.0%
Korea	How well - Presentation IC5l		0.4%	0.3%	0.3%
		1 Do well by myself	52.3%	40.8%	45.9%
		2 Do with help	31.1%	36.5%	34.1%
		3 Know but can't do	9.8%	14.2%	12.2%
		4 Don't know	6.5%	8.2%	7.4%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Download music IC5m		5.0%	5.9%	5.5%
		1 Do well by myself	60.5%	78.0%	69.2%
		2 Do with help	19.1%	10.4%	14.8%
		3 Know but can't do	13.8%	4.1%	8.9%
		4 Don't know	1.6%	1.6%	1.6%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Download music IC5m		1.3%	3.3%	2.3%
		1 Do well by myself	71.7%	84.4%	78.1%
		2 Do with help	17.9%	6.2%	12.1%
		3 Know but can't do	7.9%	3.8%	5.8%
		4 Don't know	1.2%	2.3%	1.8%
	Total		100.0%	100.0%	100.0%
Korea	How well - Download music IC5m		0.4%	0.3%	0.3%
		1 Do well by myself	94.9%	96.8%	96.0%
		2 Do with help	3.6%	2.0%	2.7%
		3 Know but can't do	0.5%	0.5%	0.5%
		4 Don't know	0.6%	0.4%	0.5%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Multi-media IC5n		4.9%	5.9%	5.4%
		1 Do well by myself	35.0%	57.1%	46.0%
		2 Do with help	42.6%	27.6%	35.1%
		3 Know but can't do	16.0%	8.5%	12.3%
		4 Don't know	1.5%	0.8%	1.2%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Multi-media IC5n		1.3%	3.3%	2.3%
		1 Do well by myself	22.7%	47.9%	35.2%
		2 Do with help	41.0%	33.9%	37.5%
		3 Know but can't do	30.3%	11.3%	20.9%
		4 Don't know	4.7%	3.6%	4.2%
	Total		100.0%	100.0%	100.0%
Korea	How well - Multi-media IC5n		0.4%	0.3%	0.3%
		1 Do well by myself	34.6%	35.7%	35.2%
		2 Do with help	42.0%	37.9%	39.7%
		3 Know but can't do	17.5%	18.7%	18.1%
		4 Don't know	5.6%	7.5%	6.7%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - E-mails IC5o		4.9%	5.9%	5.4%
		1 Do well by myself	90.6%	89.0%	89.8%
		2 Do with help	2.6%	3.7%	3.1%
		3 Know but can't do	1.2%	1.3%	1.3%
		4 Don't know	0.6%	0.2%	0.4%
	Total		100.0%	100.0%	100.0%
Iceland	How well - E-mails IC5o		1.3%	3.3%	2.3%
		1 Do well by myself	94.0%	87.7%	90.8%
		2 Do with help	3.8%	6.2%	5.0%
		3 Know but can't do	0.7%	2.2%	1.5%
		4 Don't know	0.3%	0.6%	0.4%
	Total		100.0%	100.0%	100.0%
Korea	How well - E-mails IC5o		0.4%	0.3%	.3%
		1 Do well by myself	98.6%	95.8%	97.1%
		2 Do with help	0.6%	2.1%	1.5%
		3 Know but can't do	0.1%	0.6%	0.4%
		4 Don't know	0.2%	1.2%	0.8%
	Total		100.0%	100.0%	100.0%

			TRACKING GENDER		Total
			Females	Males	1
Denmark	How well - Web Page IC5p		4.9%	5.9%	5.4%
		1 Do well by myself	24.6%	31.1%	27.8%
		2 Do with help	37.0%	41.5%	39.3%
		3 Know but can't do	31.2%	18.7%	25.0%
		4 Don't know	2.2%	2.8%	2.5%
	Total		100.0%	100.0%	100.0%
Iceland	How well - Web Page IC5p		1.3%	3.3%	2.3%
		1 Do well by myself	78.4%	56.4%	67.4%
		2 Do with help	15.7%	27.3%	21.5%
		3 Know but can't do	4.2%	10.6%	7.4%
		4 Don't know	0.4%	2.3%	1.4%
	Total		100.0%	100.0%	100.0%
Korea	How well - Web Page IC5p		0.4%	0.3%	0.3%
		1 Do well by myself	21.1%	19.2%	20.1%
		2 Do with help	49.2%	48.7%	48.9%
		3 Know but can't do	23.4%	24.6%	24.0%
		4 Don't know	5.9%	7.3%	6.7%
	Total		100.0%	100.0%	100.0%

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(98 2010 04 1 P) ISBN 978-92-64-08202-1 – No. 57417 2010

PISA Computer-Based Assessment of Student Skills in Science

This report documents the initial step towards an electronically-delivered Programme for International Student Assessment (PISA) test pioneered by Denmark, Iceland and Korea. In 2006, the PISA assessment of science included for the first time a computer-based test. The results discussed in this report highlight numerous challenges and encourage countries to take the work further.

PISA Computer-Based Assessment of Student Skills in Science describes how the 2006 survey was administered, presents 15-year-olds' achievement scores in science and explains the impact of information communication technologies on both males' and females' science skills. While males outperformed females on the computer-based test in all three countries, females in Iceland and males in Denmark performed better than their counterparts on the paper-and-pencil test. The evidence shows that, overall, males are more confident and use computers more frequently. While females tend to use the Internet more for social networking activities, males tend to browse the Internet, play games and download software.

Readers will also learn how students reacted to the electronic questionnaire and how it compared with pencil-and-paper tests. In general, there were no group differences across test methods but students enjoyed the computer-based test more than the paper-and-pencil test.

FURTHER READING

PISA 2006: Science Competencies for Tomorrow's World (OECD, 2007)

THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT (PISA)

PISA is a collaborative process among the 30 member countries of the OECD and more than 30 partner countries and economies. It brings together scientific expertise from the participating countries and is steered by their governments on the basis of shared, policy-driven interests. PISA is an unprecedented attempt to measure student achievement, as is evident from some of its features:

– *The literacy approach*: PISA aims to define each assessment area (mathematics, science, reading and problem solving) not mainly in terms of mastery of the school curriculum, but in terms of the knowledge and skills needed for full participation in society.

– *A long-term commitment*: It will enable countries to monitor regularly and predictably their progress in meeting key learning objectives.

– *The age-group covered:* By assessing 15-year-olds, *i.e.* young people near the end of their compulsory education, PISA provides a significant indication of the overall performance of school systems.

– *The relevance to lifelong learning*: PISA does not limit itself to assessing students' knowledge and skills but also asks them to report on their own motivation to learn, their beliefs about themselves and their learning strategies.

The full text of this book is available on line via these links:
www.sourceoecd.org/scienceIT/9789264082021
www.sourceoecd.org/education/9789264082021

Those with access to all OECD books on line should use this link:
www.sourceoecd.org/9789264082021

SourceOECD is the OECD online library of books, periodicals and statistical databases.
For more information about this award-winning service and free trials ask your librarian, or write to us at **SourceOECD@oecd.org**.

OECD*publishing*
www.oecd.org/publishing

ISBN 978-92-64-08202-1
98 2010 04 1 P