

**An uncomfortable dilemma –  
performance on international  
surveys and national  
assessments: the case of PIRLS  
2001 and 2006 in England**

**Liz Twist**

**National Foundation for Educational Research, UK**

---

What can international comparisons tell us about trends in national attainment?

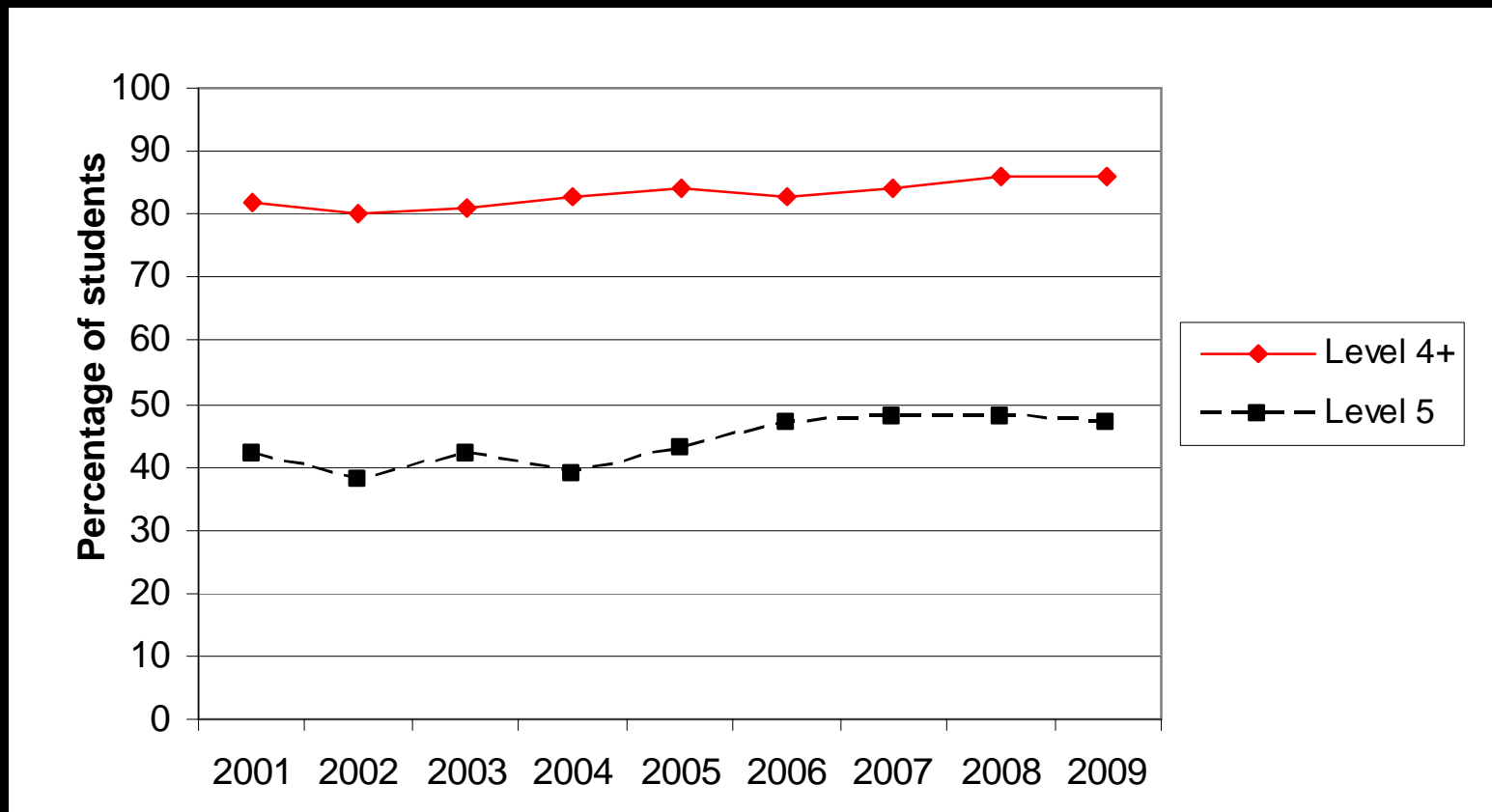
Preliminary investigation into PIRLS and data from England, the Netherlands and Sweden

# The context in England

---

- | High-stakes, high-profile national assessment system
- | Considerable investment in education
- | Concern about slow rate of progress

# Reading national test at age 11

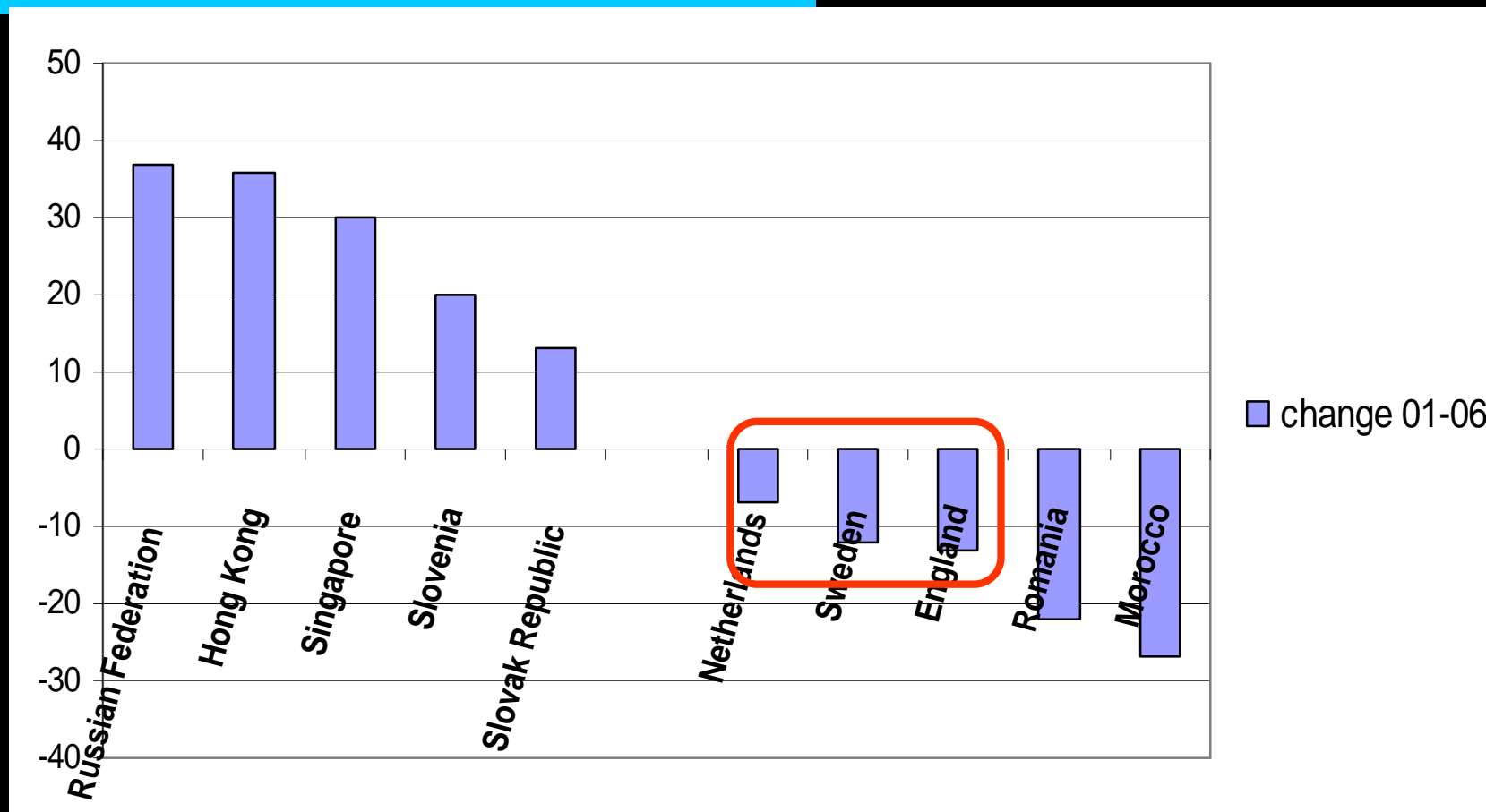


# Progress in International Reading Literacy Study (PIRLS)

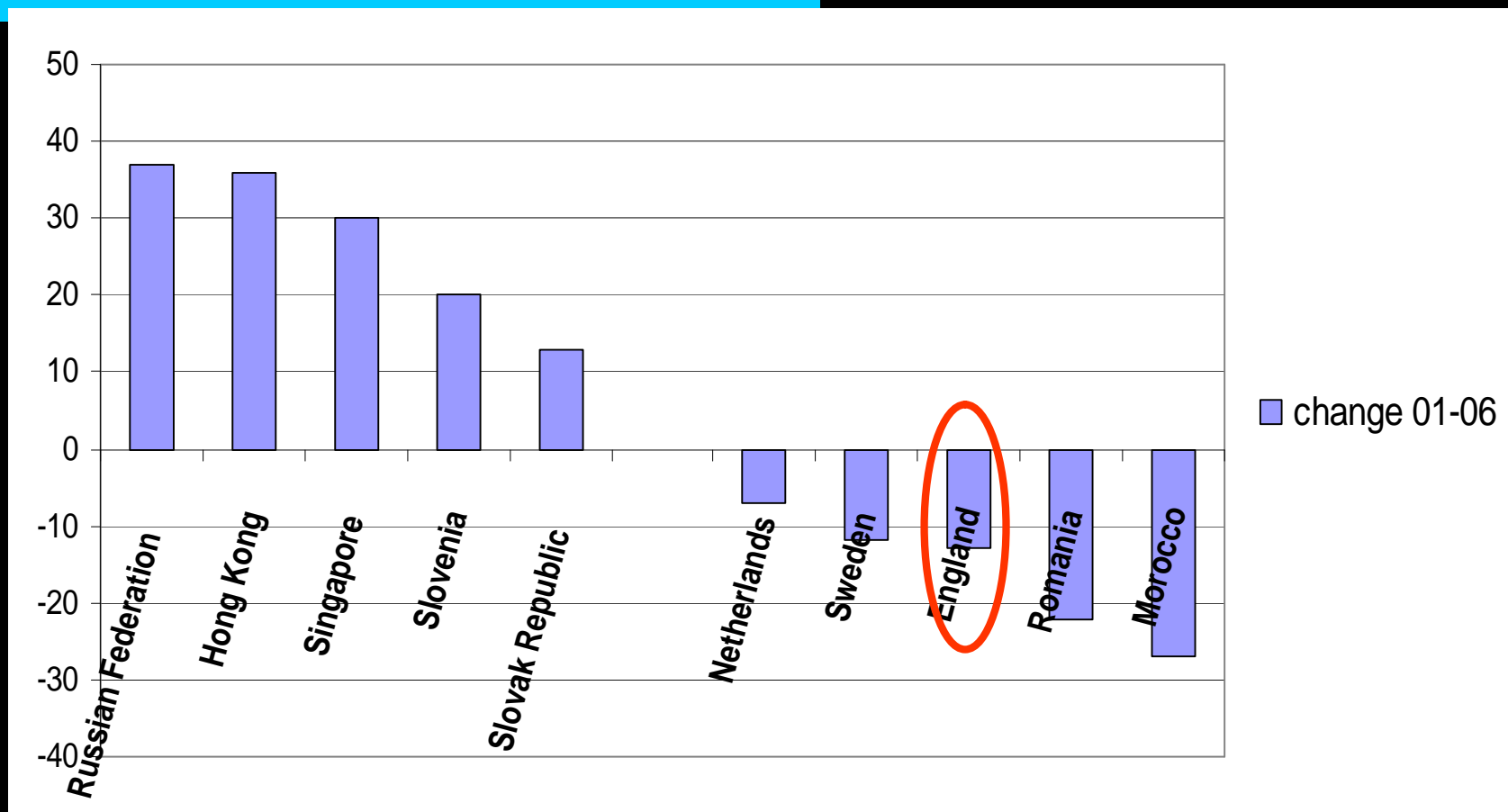
---

- | IEA
- | Reading skills of 9-10 year olds
- | Every 5 years
- | In 2001 35 countries
- | In 2006 40 countries
- | Overall reading scale, and scales for literary and information reading

# Trends in PIRLS (2001-2006)



# What our policy-makers and (some) media were interested in



What the rest of the  
media were  
interested in ...

Countries

- <sup>2a</sup> Russian Federation
- Hong Kong SAR
- <sup>2a</sup> *Canada, Alberta*
- Singapore
- <sup>2a</sup> *Canada, British Columbia*
- Luxembourg
- <sup>2a</sup> *Canada, Ontario*
- Italy
- Hungary
- Sweden
- Germany
- <sup>†</sup> Netherlands
- <sup>12a</sup> Belgium (Flemish)
- <sup>2a</sup> Bulgaria
- <sup>2a</sup> Denmark
- Canada, Nova Scotia*
- Latvia
- <sup>12a</sup> United States
- England
- Austria
- Lithuania
- Chinese Taipei
- Canada, Quebec*
- New Zealand
- Slovak Republic
- <sup>†</sup> Scotland
- France
- Slovenia
- Poland
- Spain
- <sup>2b</sup> Israel
- Iceland
- PIRLS Scale Avg.**
- Moldova, Rep. of
- Belgium (French)
- <sup>‡</sup> Norway
- Romania
- <sup>2a</sup> Georgia
- Macedonia, Rep. of
- Trinidad and Tobago
- Iran, Islamic Rep. of
- Indonesia
- Qatar
- Kuwait
- Morocco
- South Africa

# The dilemma ...

---

- | England's scale score was significantly lower in 2006 than in 2001 suggesting significantly weaker performance
- | (also evident in data from Sweden and the Netherlands)
- | But there was no fall in performance in England's national data over this period

# This led to further investigation

---

- | England, Sweden and the Netherlands had lower scale scores overall and on the two subscales
- | All three countries showed a greater fall on the literary subscale than on the information subscale
- | Back to the raw data

# Structure of the assessments

---

- | 10 passages plus associated items
- | 5 literary / 5 informational
- | 2 of each type was brought forward from 2001 – these were used for linking the two studies to produce the trend data

# England: linking blocks

<b>Block</b>	<b>2001 mean</b>	<b>2006 mean</b>	<b>Difference</b>
Literary C	10.21	10.50	+0.29
Literary F	11.59	11.07	-0.52
Information A	11.90	10.89	-1.01
Information L	9.57	8.91	-0.36

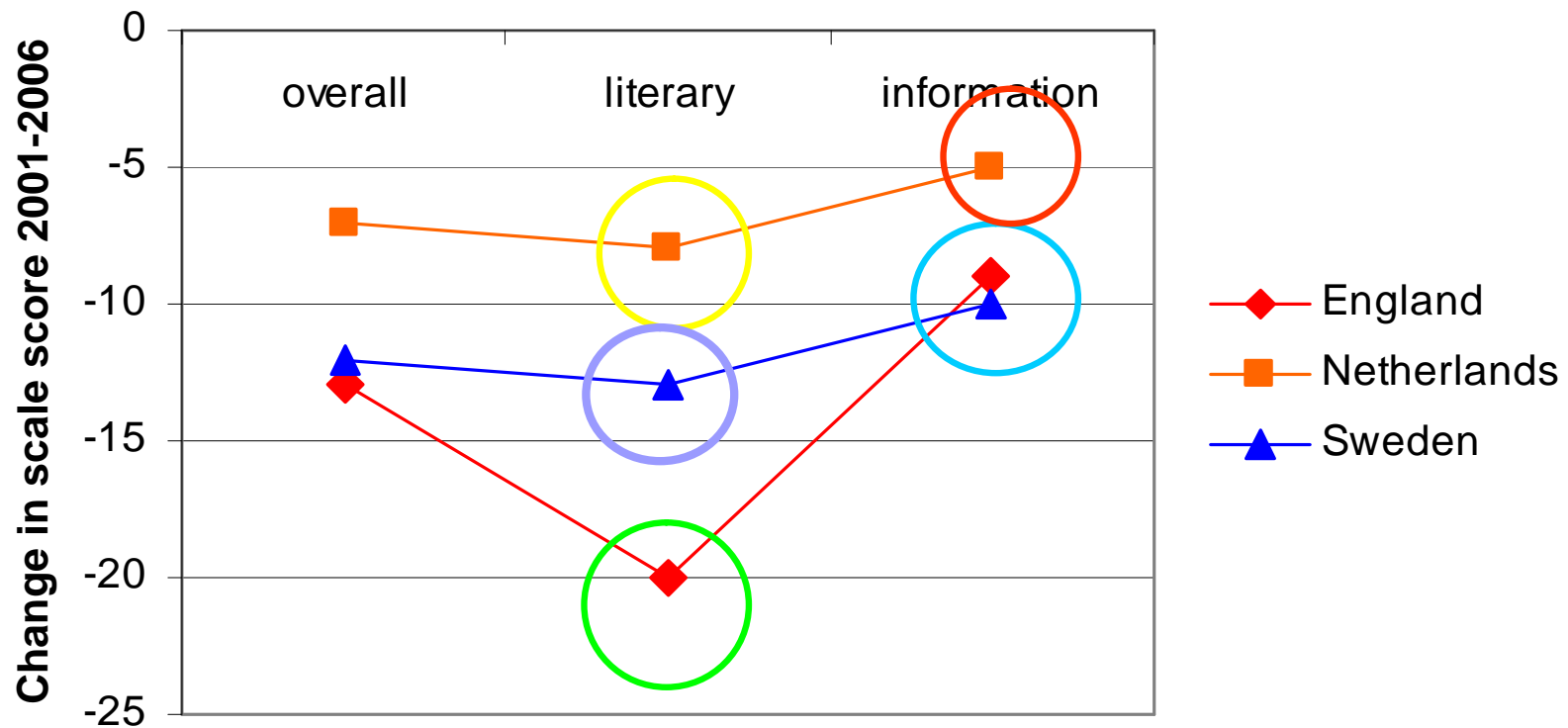
# Netherlands: linking blocks

<b>Block</b>	<b>2001 mean</b>	<b>2006 mean</b>	<b>Difference</b>
Literary C	10.65	11.09	+0.44
Literary F	11.26	11.55	+0.29
Information A	12.30	12.01	-0.29
Information L	8.89	9.23	+0.34

# Sweden: linking blocks

<b>Block</b>	<b>2001 mean</b>	<b>2006 mean</b>	<b>Difference</b>
Literary C	11.26	10.94	-0.32
Literary F	12.14	11.70	-0.44
Information A	12.30	12.11	-0.19
Information L	8.80	9.05	+0.25

# Summary of trend outcomes



# The puzzle ...

---

In the trend data in the international report, all three countries fell on the overall scale and on both subscales.

For all three countries, the fall on the literary scale was the greater.

Looking only at the linking blocks, a different picture emerges: England falling more on information, NL increasing on literary, falling on information, Sweden falling on literary, stable on information.

# The investigation ...

---

- | Currently item parameters (difficulties) are estimated using data from all countries
- | This produces an international estimate of item difficulty for each item
- | Gebhardt and Adams (2007) argue that this ignores item-by-country interactions

- 
- | What happens when linking is done for each country separately using a 2-parameter IRT model and common items?
  - | Results are closer to those from simple analysis of linking items i.e. for England, fall is apparent on information subscale, with a very slight increase on the literary subscale.

# England: 2001-2006

<b>Scale</b>	<b>2001 mean</b>	<b>2006 mean</b>	<b>Difference</b>
Literary	558.0	559.4	+1.4
Information	546.7	540.4	-6.3
Overall	552.9	548.1	-4.8

# Summary

- | When each country's data is analysed separately using 2-parameter IRT model and common items
  - For England results on literary blocks are very comparable 2001 and 2006
  - England's results on information blocks show a slight fall
  - For Sweden and the Netherlands, very little apparent change between 2001 and 2006 on either subscale

# Conclusion

---

- | Results appear sensitive to the linking methodology
- | Is this due to some countries showing marked improvement in performance in 2006 which has affected item parameters?

# For further information

---

Liz Twist

**NFER**

National Foundation for  
Educational Research

[www.nfer.ac.uk](http://www.nfer.ac.uk)