

**ITC 2010 conference in Hong Kong,**  
**Symposium: Assessment models for monitoring learning**



# **National Tests in Denmark**

## **– CAT as a Pedagogic Tool**

**Jakob Wandall,**  
**Chief Adviser,**  
**Skolestyrelsen (Danish National School Agency)**



# The Danish National testing system

- Background
- The Danish testing system
  - Design of the testing system
  - Adaptive vs. linear testing
  - Status for the project
- Challenges



# Reference countries to Denmark



# Background for the Danish initiatives to improve the culture of evaluation

## Facts of the public primary and lower secondary education in Denmark, "Folkeskolen":

- 600.000 pupils, 60.000 pr. form
- 1.300 ordinary public schools, 600 other institutions (e.g. for special need education)
- Schools are run by local governments → significant differences from municipality to municipality
- 4-5 pupils pr. new computer (<3 year)
- highspeed internet-connection almost everywhere
- **Infrastructure suited for It-based assessment**

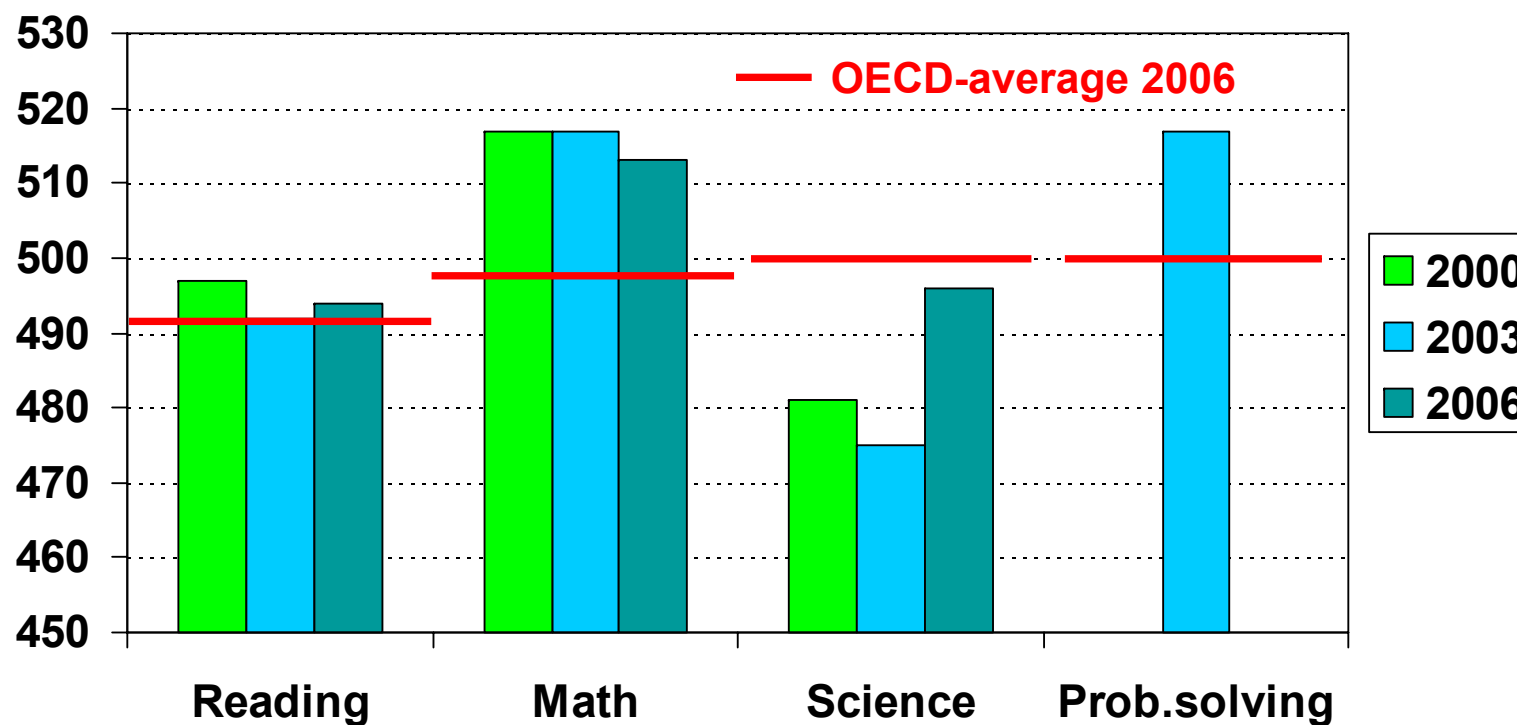


## **Background for the Danish initiatives to improve the culture of evaluation**

1. PISA-surveys (2000-2006)
2. OECD-review on Denmark (2004) and National reports on assessment in school (2004)
3. Governments initiatives according to the OECD-Recommendations



# The Danish PISA results, 2000 - 2006



Review of National Policies for Education, Denmark, OECD 2004

” Denmark has one of the most expensive education systems in the world, and for years perceived it to be one of the best in the world.

However the disappointing results of recent international tests to measure schooling outcomes confirmed earlier evidence, that the system actually is underperforming. ”



# OECD-review 2004

## Strengths, weaknesses, 35 recommendations

### Some weaknesses:

- Poor tradition of pupils assessment,
- Insufficient teacher qualifications in assessment techniques
- Insufficient exchange of best practices between teachers

### Some recommendations for the Minister:

- Development of criteria-based tests.
- Development of different assessment methods and - materials.
- Carry out a policy based on the principle, that test results don't get published for ranking purposes.



# Implementation of OECD's recommendations

- The Government initiatives followed the OECD recommendations with some modifications – e.g. in the testing system
- The OECD-team recommended criteria-based test - recommendation based on a different tradition of testing/evaluation
- Background: The OECD-team came from England, Canada and Finland



# Two different traditions for assessment/use of test results

## 1. The Nordic / continental

Terminology, Nordic/German origin (Danish: Prøve, vurdering, bedømmelse, opgave, karakter)

## 2. The Anglo-American

Terminology, English origin (Danish: test, score, item)



# Traditions for assessment/testing,

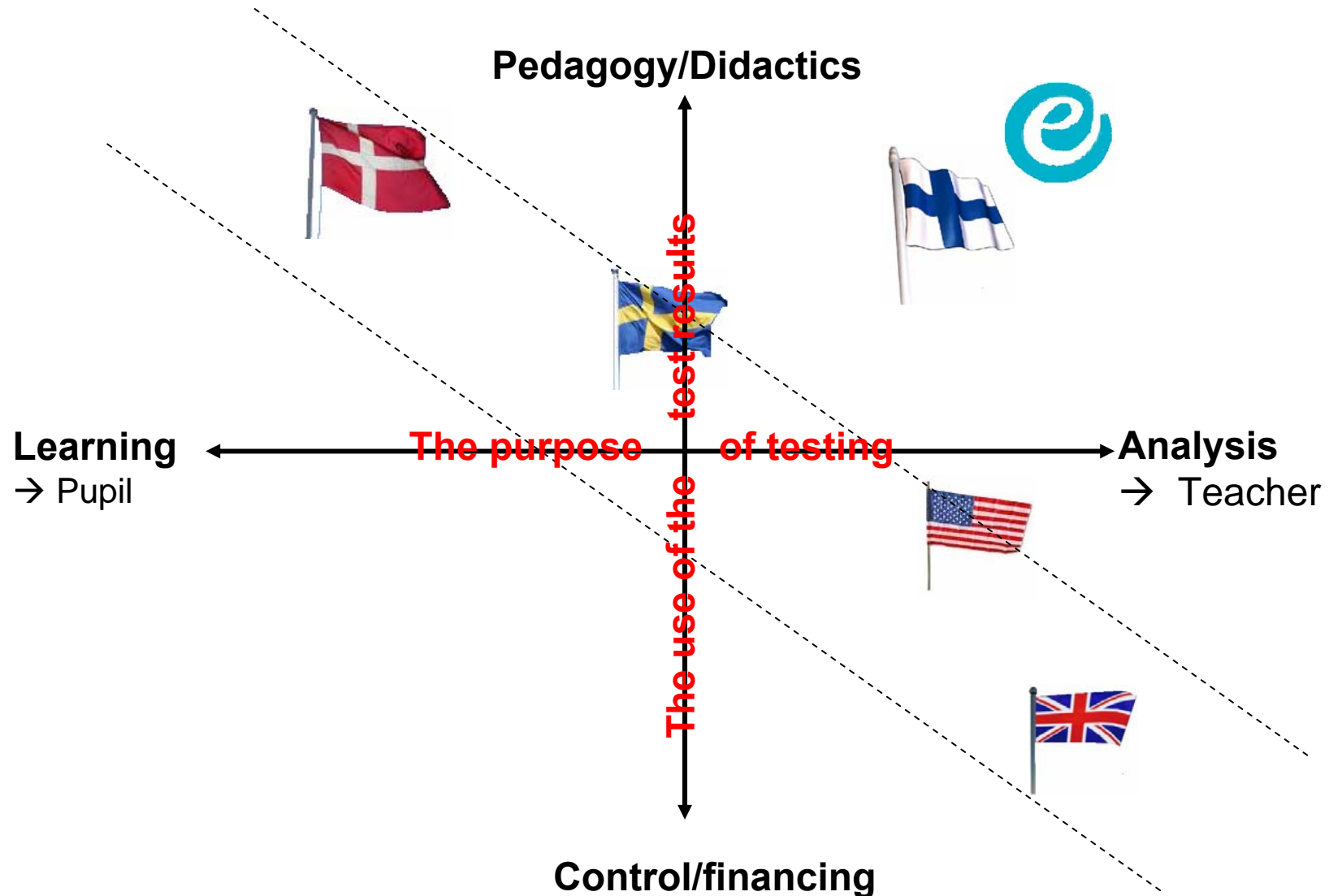
Why testing? What are the results used for?

<b>Traditions:</b>	<b>Nordic/continental</b>	<b>English/American</b>
<b>Purpose</b>	<b>Learning → Pupil</b>	<b>Analysis → The “system”/Professionals</b>
	<b>Focus on equality and solidarity</b>	<b>Focus on ambitions and elite</b>
	<b>Solicitude</b>	<b>Fairness</b>
<b>Use of results</b>	<b>Pedagogy</b>	<b>Control of outcome</b>
	<b>Didactics / Teaching</b>	<b>Financing / grants</b>
	<b>Formative / Low-stake</b>	<b>Summative / High-stake</b>



# Traditions for testing,

Why test pupils – What are the results used for



# Criteria for choice of strategy

- **Test in combination with other assessment tools**
    - Easy to use for the teacher
    - Flexible systems from the schools point of view
    - Low/no cost for the school
  - **Priority to pedagogical purposes – Formative – Low stake**
    - For teachers assessment of the pupils
    - The teacher sets the rules
  - **Effective Self-correcting tests**
    - valid,
    - reliable
    - detailed results
    - **Max. 45 minutes** (1 lesson)
- Central administered internet-based Computer adaptive testing - **with focus on pedagogy**



# The Danish testing system



# The national tests

Form	1	2	3	4	5	6	7	8	9
<b>Subject</b>									
Danish / reading		X		X		X		X	
Math			X			X			
English							X		
Geography								X	
Biology								X	
Physics/chemistry								X	
Danish as 2'nd language									



Forms where tests can be used



10 compulsory tests

# Main features of the Danish national tests

- IT-based, automatically scored, provided freely by the National School Agency.
- Pedagogical purpose → The teacher sets the rules for the test
- The teacher are booking & administering the tests and interpreting the test results
- Different feedback on various levels (teacher, headmaster, municipality)
- The teacher gets detailed results online for
  - the class,
  - the individual pupil and
  - Details of the tests, including the items in the individual tests.



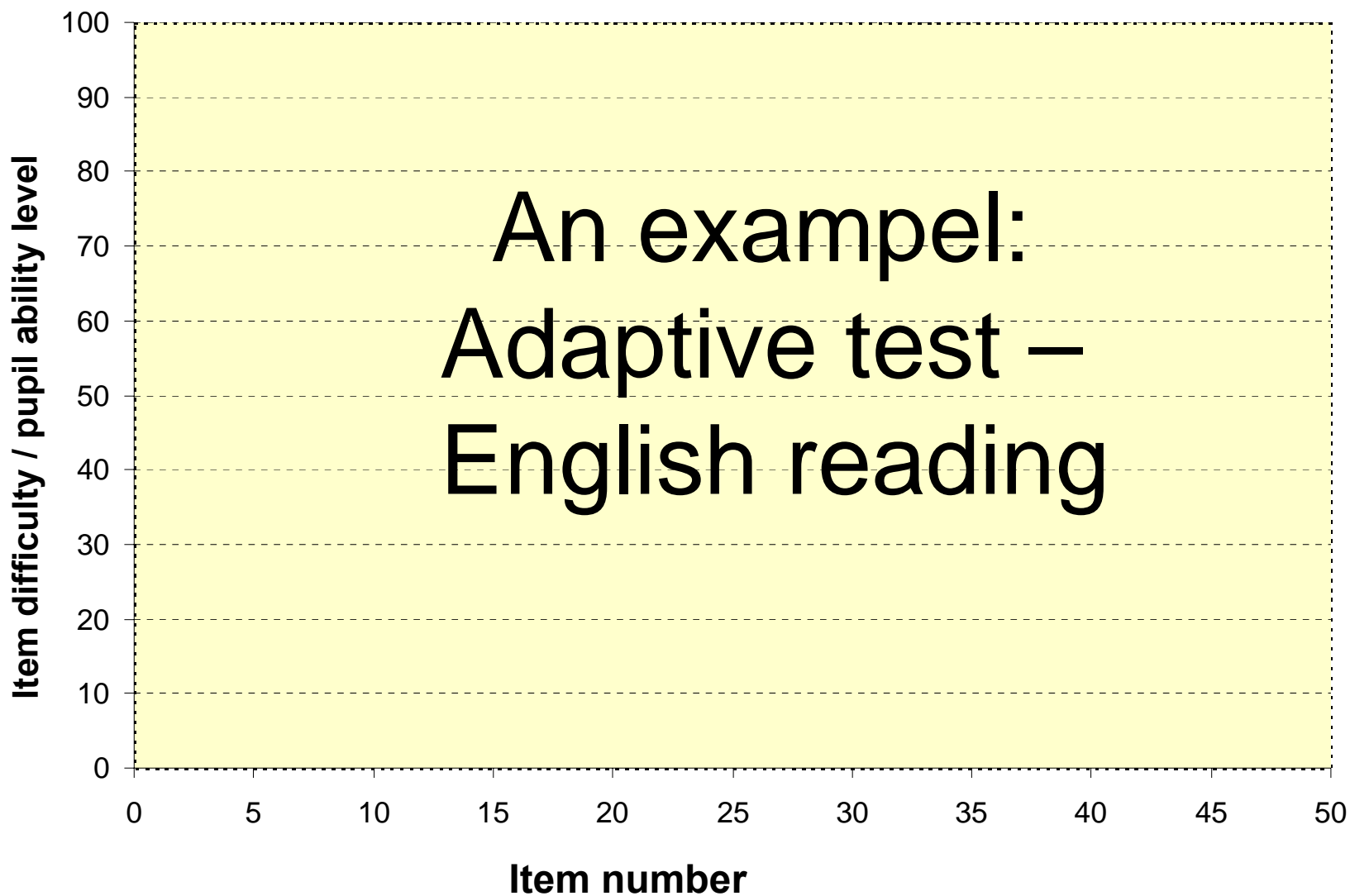
# CAT (Computer Adaptive Testing)

- Adaptive = Adapts to the individual pupils ability:  
Correct answer → More difficult questions  
Wrong answer → Easier questions

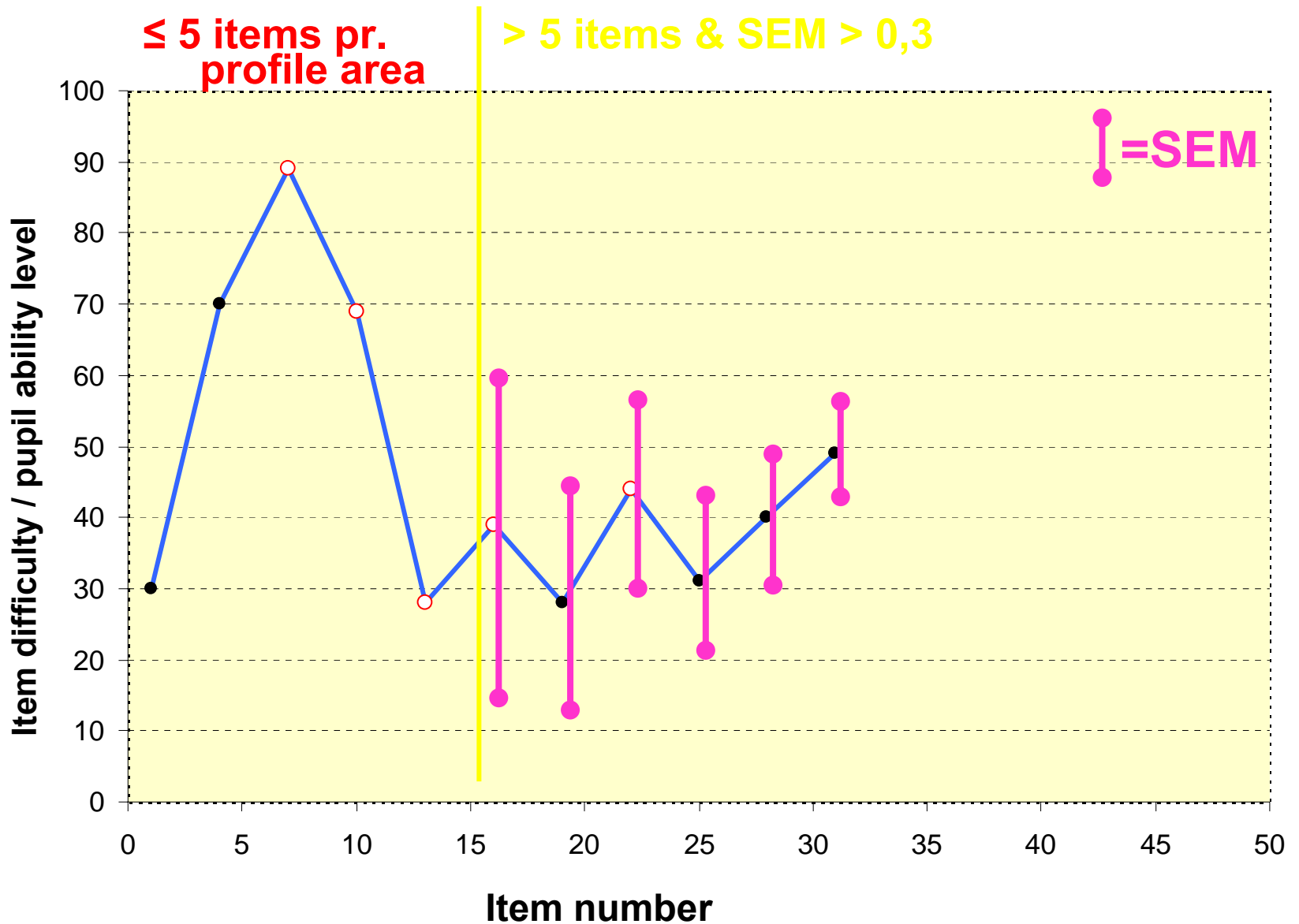
**The test is most efficient when item difficulty = pupil ability**

- More effective testing → more detailed results:  
Adaptive within 3 “profile areas” → 3 tests in 1  
(E.g. English: Reading, Vocabulary and Language usage)
- Simple principles, but a few tricky conditions:
  - Extensive demands to the technology – both capacity and stability.
  - Very large Item-pools with the exact right mix of high quality items

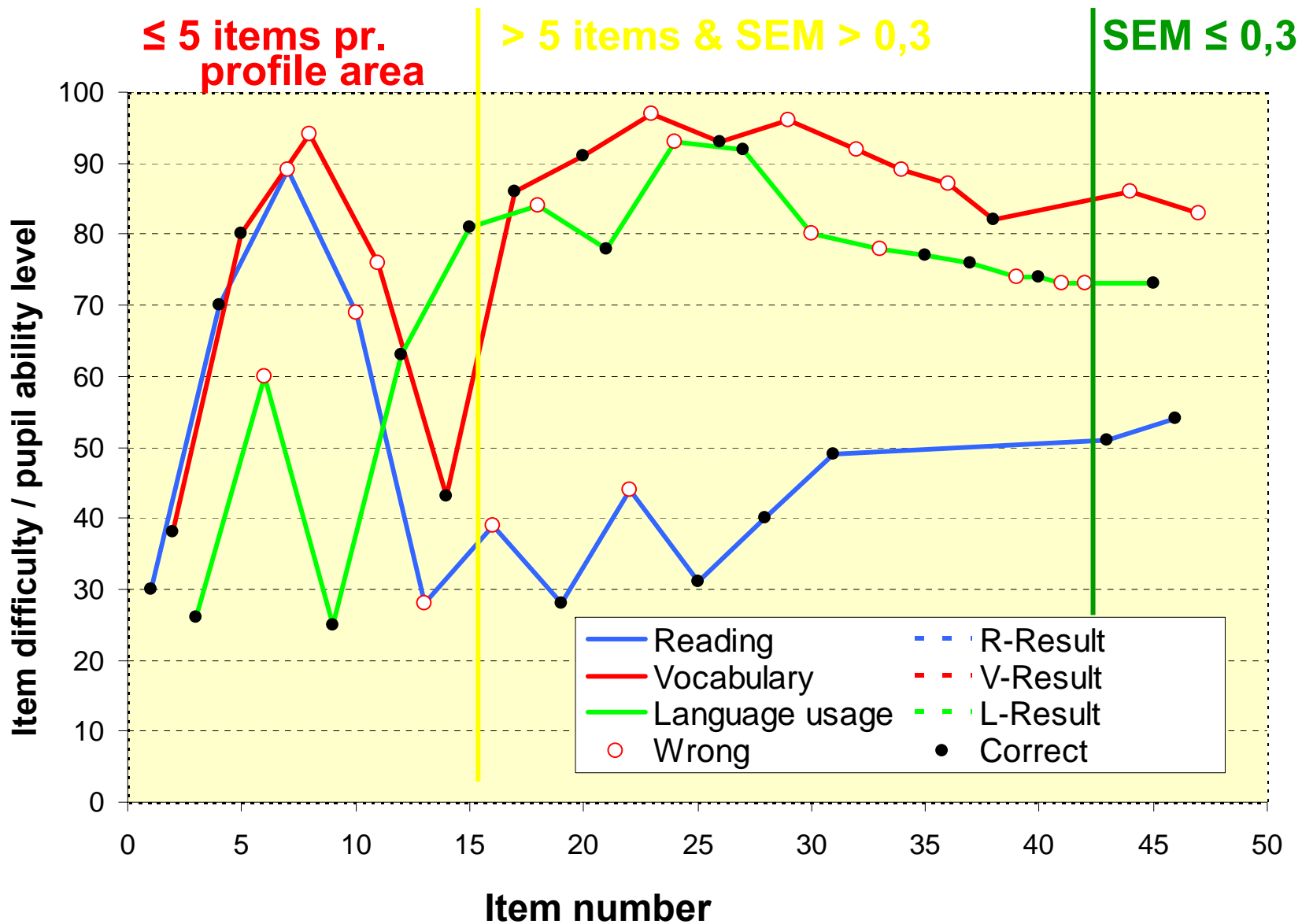




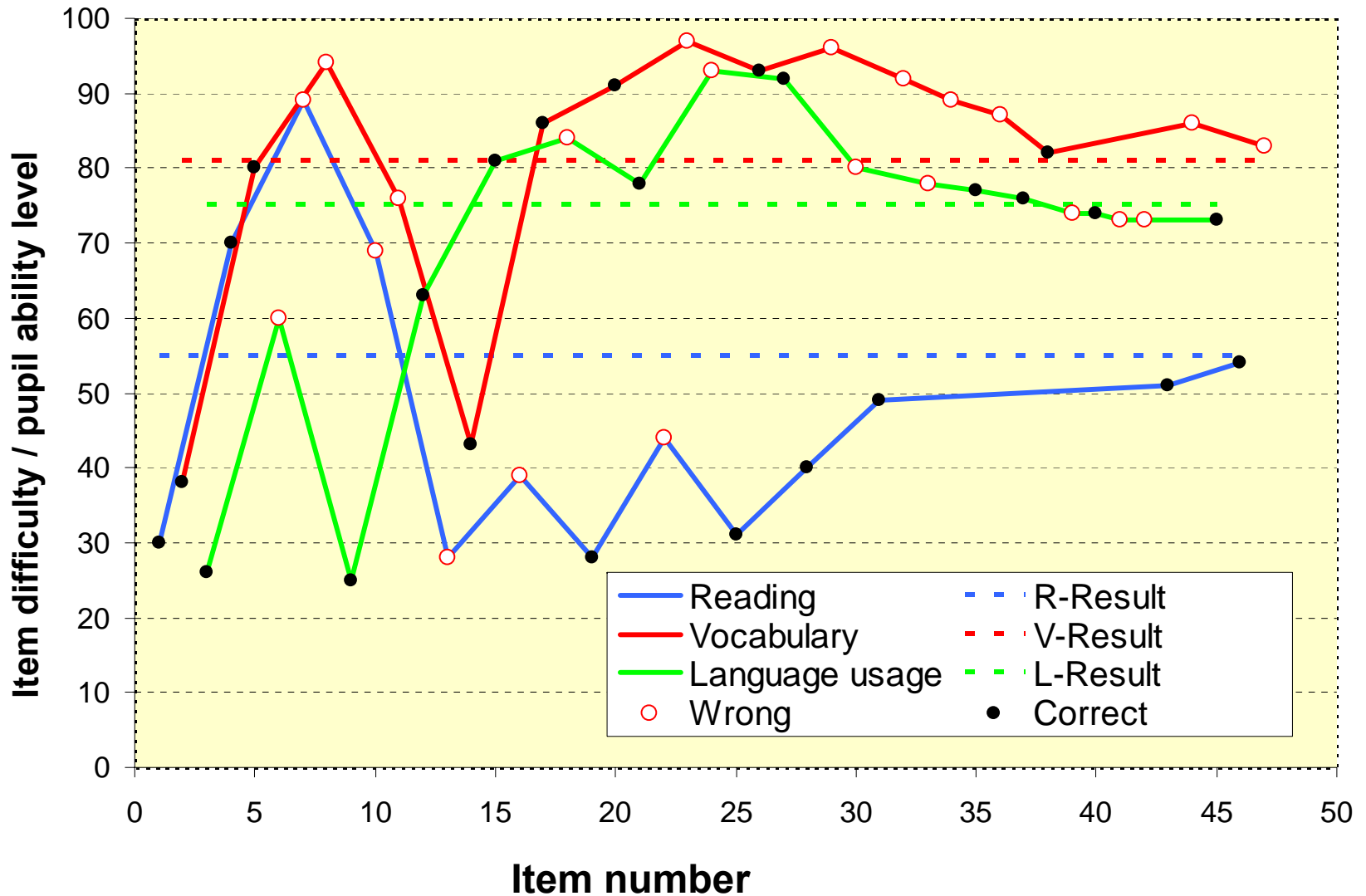
# Adaptive test – English reading



# Adaptive test – 3 profile areas



# Adaptive test – 3 profile areas

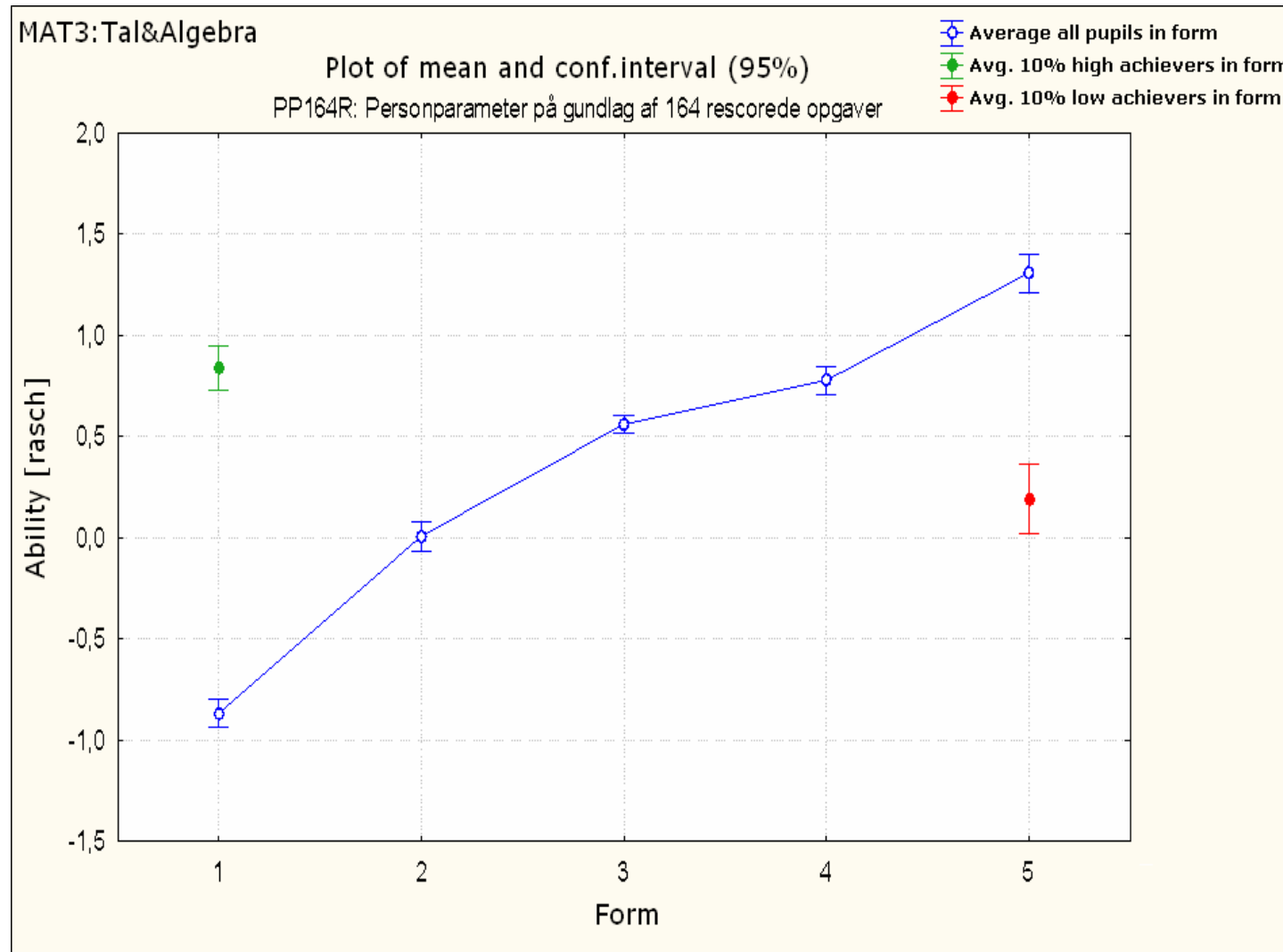


# Linear vs adaptive testing

- The adaptive model performs relatively best in the extremes (measuring proficiency of high and low achievers).
- The adaptive setup is pretty efficient: Reduction of SEM to 0,3 normally within 15 items.
- The adaptive test is generally a good experience for the low achievers, but can be a frustrating experience for high achievers (no free rides).
- The greater the differences in ability within the tested group of pupils, the greater the advantage of adaptive testing.
- A comprehensive school with significant academic differences in every classroom

# Danish experiences: linear vs adaptive

Progression and differences between pupils in the same form



# Different types of questions/items

- Both dichotomous and polytomous items are used
- Assessment of item response based on a partial credit model
- Polytomous items/partial credit provides new opportunities of adaptive testing (within the Rasch model). E.g. we are working on inclusion of the time dimension in the estimation of the pupils reading ability.



# Controlling the items

Condition for adaptive testing is:

- That the difficulty of the items are well defined and stable over time (homogeneity).
- No differential item function.
- Sufficient number of items, so that even the fastest pupils don't run out of items
- That the items are evenly distributed on item-difficulty  
→ challenges for all pupils.

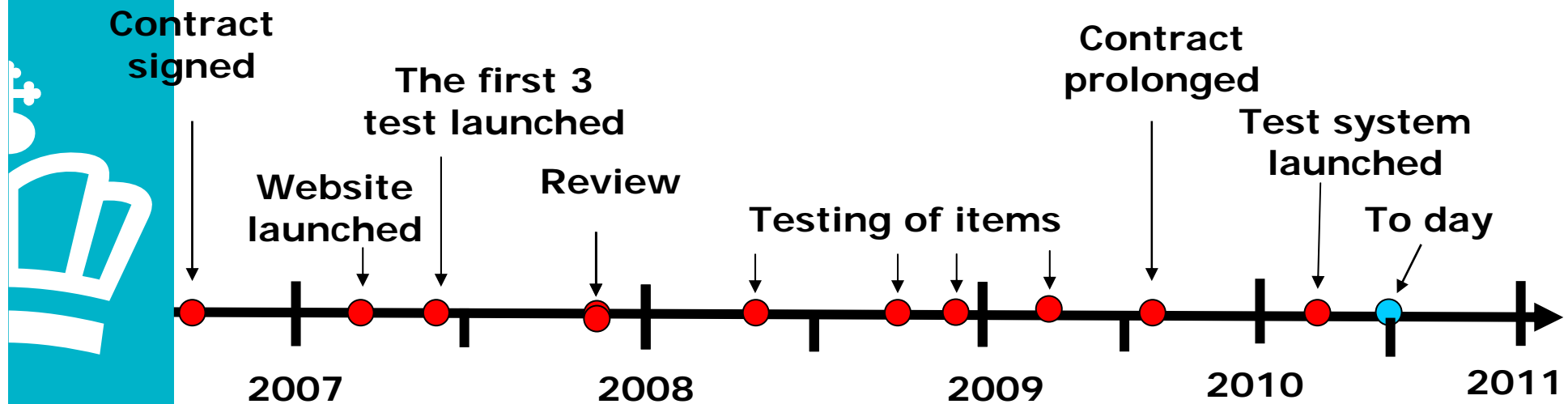
Therefore following requirements have to be met:

- Minimum 540 items pr test (180 pr. Profile area) – evenly distributed on difficulty levels
- All items are tested on a large number of pupils (500-700)
- The items are required to fit a Rasch model
- Not more than 3 runs pr pupil/test (including compulsory test)



# Status on the implementation

- The first 3 test was launched 2007, with reduced item banks
- There has been conducted an expert review that showed, that
  - the psychometric standard was at a very high level, but
  - the quality of the items and size of the item pools were insufficient
- Development and testing of items in 2008-2009
- Full scale tests of the test system in autumn 2009
- Test system launched march 2010



# Challenges for Denmark

- To make the it-systems work flawlessly (booking, test and results)
- To further develop the test system (e.g. only one item pool pr. subject, including time dimension in the difficulty - decoding items)
- To maintain and further develop high quality item pools
- To develop user friendly ways to present the results
- To convince the teachers/Schools to use the information – and describe how to do it (guidance in good practice) !!





Alternatives to  
external standardized assessments

Liz Hamp-Lyons

# Overview

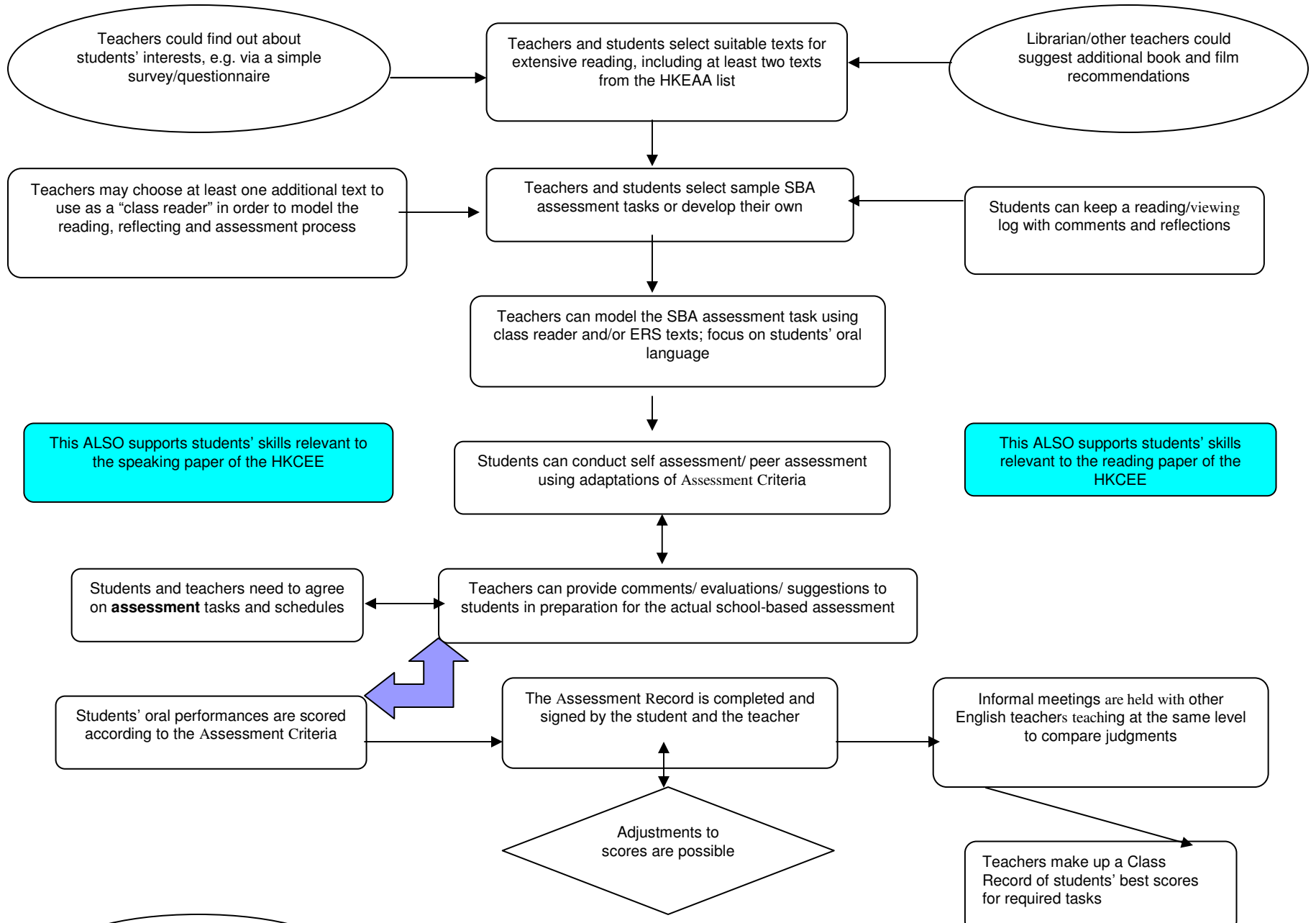


- 1) description of English oral SBA
  - HK Certificate of Education
  - HK Diploma of Education
- 2) how and to what extent this school-based assessment:
  - supports learning in the classroom
  - contributes to teacher professional development
- 3) use of teacher-assessed performance scores for educational planning and region-wide accountability



# What happens in the school

- Every teacher carries responsibility for SBA in her own classroom
- Every school's English Panel Chair ("Head of English") carries responsibility for monitoring within-class quality of teaching, and for between-class parity of standards for assessing SBA performances
- Every school's Principal carries responsibility to ensure SBA is implemented according to HKEAA regulations and SBA principles





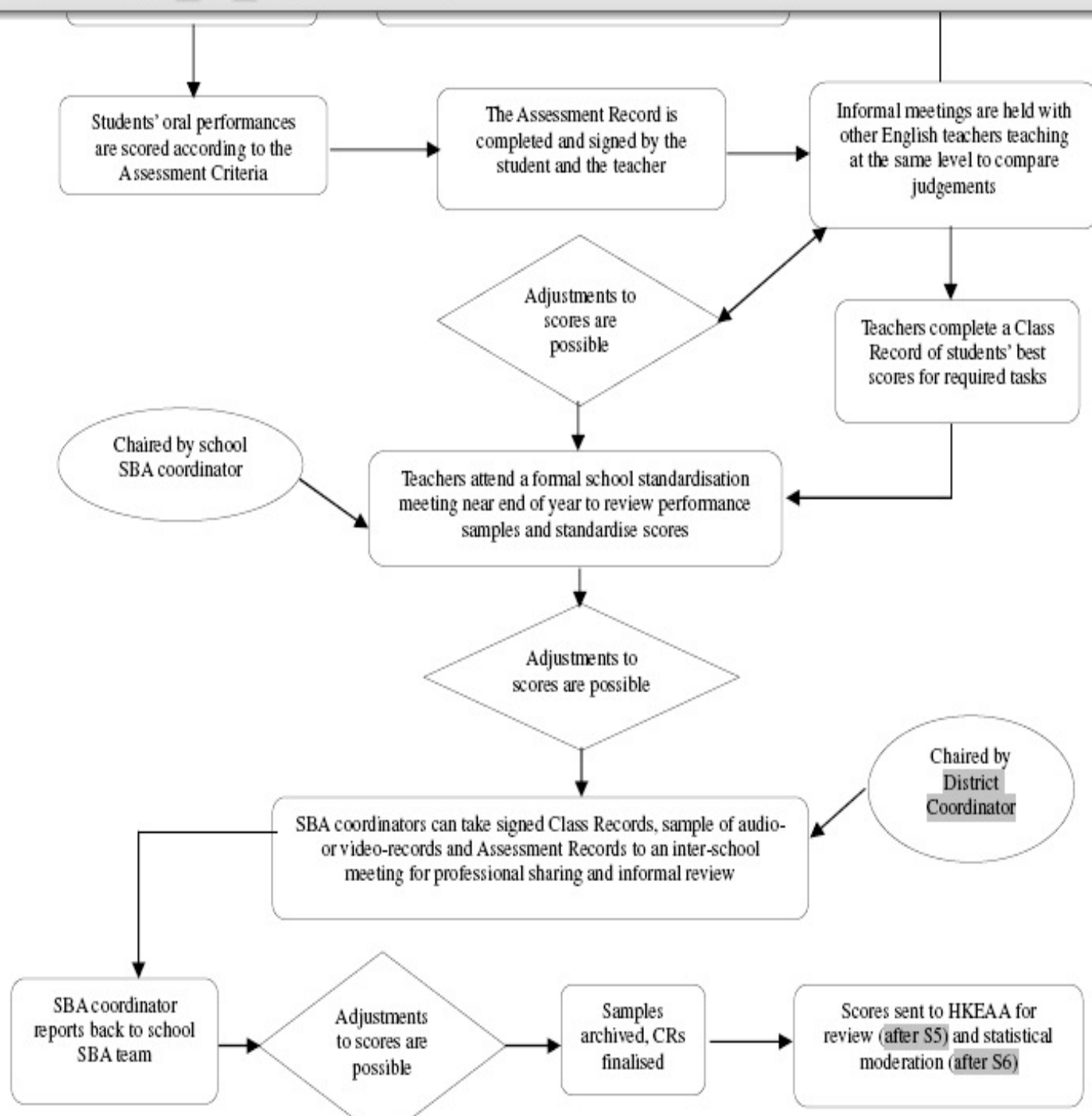
# What assessment based with the teacher/in the school demands

- Support within the school
- Support within the Hong Kong society
- Ongoing assessment tools renewal
- Ongoing curriculum renewal
- **Teacher professional development**
- Ongoing PD materials renewal
- Available refresher and advanced PD opportunities for teachers



# What happens after the school

- Inter-school moderation-
  - Approx.45 District Coordinators each advising 10-15 schools
  - Oversight of SBA process by HKEAA Senior Subject Officer, English
- Territory-wide statistical moderation by HKEAA





# Statistical Moderation of SBA Marks

SBA is only 15% of a traditional schools exam:

English Language results reported:

- Reading
- Writing
- Listening and integrated skills
- Speaking, including SBA performance

No of English exam candidates:	2007	99,771
	2008	101,656



# Inter-paper correlations 2007

Correlation Matrix (Option 1 and all papers sat, N=28,253)

	<b>R</b>	<b>W</b>	<b>L &amp; IS</b>	<b>S</b>	<b>SBA</b>
<b>R</b>	1.000	.858	.887	.776	.803
<b>W</b>	.858	1.000	.852	.767	.797
<b>L &amp; IS</b>	.887	.852	1.000	.764	.796
<b>S</b>	.776	.767	.764	1.000	.787
<b>SBA</b>	.803	.797	.796	.787	1.000

# Mean SBA Scores for Option 1 schools\*

The mean of the SBA scores is ...	No. of Schools	Percentage (%)
within the range	144	72.4
slightly higher than expected	29	14.6
higher than expected	2	1.0
much higher than expected	0	0
slightly lower than expected	21	10.6
lower than expected	3	1.5
much lower than expected	0	0

\* Lee, HKEAA, IAEA 2008, reporting on SBA 2007

# Inter-paper correlations 2008

Correlation Matrix (Option 1 and all papers sat, N=40,762)

	<b>R</b>	<b>W</b>	<b>L &amp; IS</b>	<b>S</b>	<b>SBA</b>
<b>R</b>	1.000	.850	.888	.759	.781
<b>W</b>	.850	1.000	.871	.773	.802
<b>L &amp; IS</b>	.888	.871	1.000	.782	.808
<b>S</b>	.759	.773	.782	1.000	.784
<b>SBA</b>	.781	.802	.808	.784	1.000



## Mean SBA Scores for Option 1 schools\*

<b>The mean of the SBA scores is ...</b>	<b>No. of schools</b>	<b>Percentage</b>
<b>within the range</b>	<b>203</b>	<b>71.5</b>
<b>slightly higher than expected</b>	<b>36</b>	<b>12.7</b>
<b>higher than expected</b>	<b>7</b>	<b>2.5</b>
<b>much higher than expected</b>	<b>2</b>	<b>0.7</b>
<b>slightly lower than expected</b>	<b>32</b>	<b>11.3</b>
<b>lower than expected</b>	<b>4</b>	<b>1.4</b>
<b>much lower than expected</b>	<b>0</b>	<b>0.0</b>

\* Chan, HKEAA SBA Report 2008



# What are the effects of moderation on students?

## **2007:**

79% of students' scores were adjusted by 0 – 3 marks  
[<1% of paper]

1% of students' scores were adjusted by 7-9 marks  
[<3% of paper]

## **2008:**

82% of students' scores were adjusted by 0-3 marks  
[<1% of paper]

2% of students' scores were adjusted by 7-9 marks  
[<3% of paper]



- For information about the HKEAA method of statistical moderation, you can download the booklet (bilingual Chinese and English) at:  
[www.hkeaa.edu.hk/DocLibrary/SBA/booklet\\_sba.pdf](http://www.hkeaa.edu.hk/DocLibrary/SBA/booklet_sba.pdf)



# Willingness to try

- SBA takes place in school at the school's convenience
- Written papers are taken in May
- The formal Speaking exam (2-rater OPI) is taken in June
- Candidates who feel they did badly in earlier exams tend to absent themselves from the Speaking exam
- **There is some evidence that SBA has had a positive effect on attendance rates for the Speaking exam**



# Just showing up...

- In 2007, 9.3% of Option 1 and 2 school SBA students failed to show up; 16% of Option 3 schools failed to show up
- In 2008, the HKEAA SBA Supervisor said in her Report:
- “Since its implementation in September 2005, both teachers and students have become more familiar with SBA and this has raised their confidence levels. The students are more confident in speaking in English and this has been reflected in the number of students who are more willing to speak during the public examination and in the lower absentee rate for the latter.”

# Indicators of beneficial consequences



The beneficial washback of SBA can be seen in the results of the speaking examination. Candidates from schools that submitted SBA marks had a lower absentee rate of 6% compared to 16% for candidates from schools that chose not to submit SBA marks. Candidates who did SBA in school also performed better in the speaking examination than those from schools not submitting SBA marks.

Christina Lee, Senior English Subject Officer, HKEAA  
(Reporting on the 2007 results)  
IAEA Conference 2008



# Longitudinal research

- Liz Hamp-Lyons and Chris Davison have been leading a 4-year study of a cross-section of schools to examine the impact of schools' choices of when to implement SBA; teachers' confidence or otherwise with SBA; their attitudes to SBA over time; their sense of whether their students are being helped or hindered by SBA; and any specific issues within the school.

# Thank you

## ACKNOWLEDGEMENTS:

Funding for development of the SBA English oral for HKCEE and HKDSE, came from the Hong Kong Education and Examinations Authority, which also continues to fund the development and delivery of teacher professional development by DVD, online and in face-to-face workshop series

TO VIEW PROFESSIONAL DEVELOPMENT MATERIALS visit <http://web.hku.hk/~sbapro>

LIZ HAMP-LYONS:

[Liz.hamp-lyons@nottingham.ac.uk](mailto:Liz.hamp-lyons@nottingham.ac.uk)

[lizhl@hkucc.hku.hk](mailto:lizhl@hkucc.hku.hk)





# Assessment Criteria

## Four domains of assessment:

- **Pronunciation & Delivery**

Phonology & intonation, voice projection & fluency

- **Communication Strategies**

Body language, timing, asking/responding to questions

- **Vocabulary & Language Patterns**

Vocabulary, language patterns, & self-correction/reformulation

- **Ideas & Organization**

Expressing information and ideas, elaboration on appropriate

topic, organization, questioning & responding to questions





# ASSESSMENTS FOR MONITORING AND IMPROVING LEARNING

Kadriye Ercikan  
University of British Columbia  
Canada  
[Kadriye.ercikan@ubc.ca](mailto:Kadriye.ercikan@ubc.ca)

# Dissatisfaction with Large-scale Assessments

- Growing during the last 15 years
- Race to The Top legislation
- Foundation Skills Assessments and Vancouver School Board
- Presentations about Denmark, Hong Kong and New Zealand Assessments

# Outline

- Canadian context
- Ercikan and McKeown Guidelines for large-scale assessments
- Similarities and differences among the Denmark, Hong Kong and New Zealand assessments
- Further insights about the role of classroom based assessments in large-scale assessments

# The Canadian Context

- Large-scale assessments inform decision making at school, district and provincial levels
- Problems with school rankings
- Teachers protesting assessments

# Ercikan and McKeown Guidelines (2007)

- Assessments need to assess valued outcomes
- Assessments need to be aligned with learning, instructional and curricular goals
- Assessments need to provide accurate estimates of student knowledge and competencies
- Score reports need to be informative and developed in a timely manner; and
- Consequential aspects of assessment results need to be an important consideration.

# Assessments need to assess valued outcomes

<b>Denmark Assessment (Wandall, 2008)</b>	<b>Hong Kong (Hamp-Lyons, 2009)</b>	<b>New Zealand (Hattie, 2006)</b>
Assessment can handle a variety of item types Publishing test results is forbidden by law	1. Foster student learning and development	1. National Assessment should mirror what is important 2. Make the “rich ideas” underlying the curricula not the items dominant

# Assessments need to be aligned with learning and instructional and curricular goals

<b>Denmark Assessment (Wandall, 2008)</b>	<b>Hong Kong (Hamp-Lyons, 2009)</b>	<b>New Zealand (Hattie, 2006)</b>
Assessment embedded within instruction	1. Foster student learning and development	7. Assess what has been taught, not teach what is to be assessed
Teachers have control over when assessments are administered	2. Construct Validity	

# Assessments need to provide accurate estimates of student knowledge and competencies

<b>Denmark Assessment (Wandall, 2008)</b>	<b>Hong Kong (Hamp-Lyons, 2009)</b>	<b>New Zealand (Hattie, 2006)</b>
CAT optimizes measurement accuracy		

# Score reports need to be informative and provided in a timely manner

<b>Denmark Assessment (Wandall, 2008)</b>	<b>Hong Kong (Hamp-Lyons, 2009)</b>	<b>New Zealand (Hattie, 2006)</b>
Score reports and feedback are immediately available and are detailed at sub-content area level		8. Ensure the assessment system provides feedback information to teachers, students, principals, systems, and parents about

# Consequential aspects of assessment results need to be an important consideration

<b>Denmark Assessment (Wandall, 2008)</b>	<b>Hong Kong (Hamp-Lyons, 2009)</b>	<b>New Zealand (Hattie, 2006)</b>
These assessments are formative and low stakes	3. Fostering an atmosphere of fairness  4. Fostering educational improvement	3. Reduce the stakes 4. Make 'National' Assessment part of a larger strategy 5. Ensure comparability but avoid league tables 6. Ensure that the assessments are valued by teachers

# Concluding comments

- Classroom-based assessments essential components of large-scale assessments for improving learning.
- Different teacher engagement with assessment.

# Denmark

- Teachers choose when tests are administered but tests are optimally selected by the computer for each individual student.
- Data are collected ten times over the academic year and scores and feedback are provided to students and teachers immediately.

# Hong Kong

- Teachers administer components of the assessments that allow interaction between teachers and students (reading, viewing and talking about texts).
- Teachers score their own students and can provide immediate feedback to students

# New Zealand

- Teachers develop tests that fit their instructional goals in a psychometrically optimal way.
- The score reports are designed to guide teachers instruction and feedback to students.