

Embedded fieldtesting in large scale tests

Frans Kleintjes



now you know



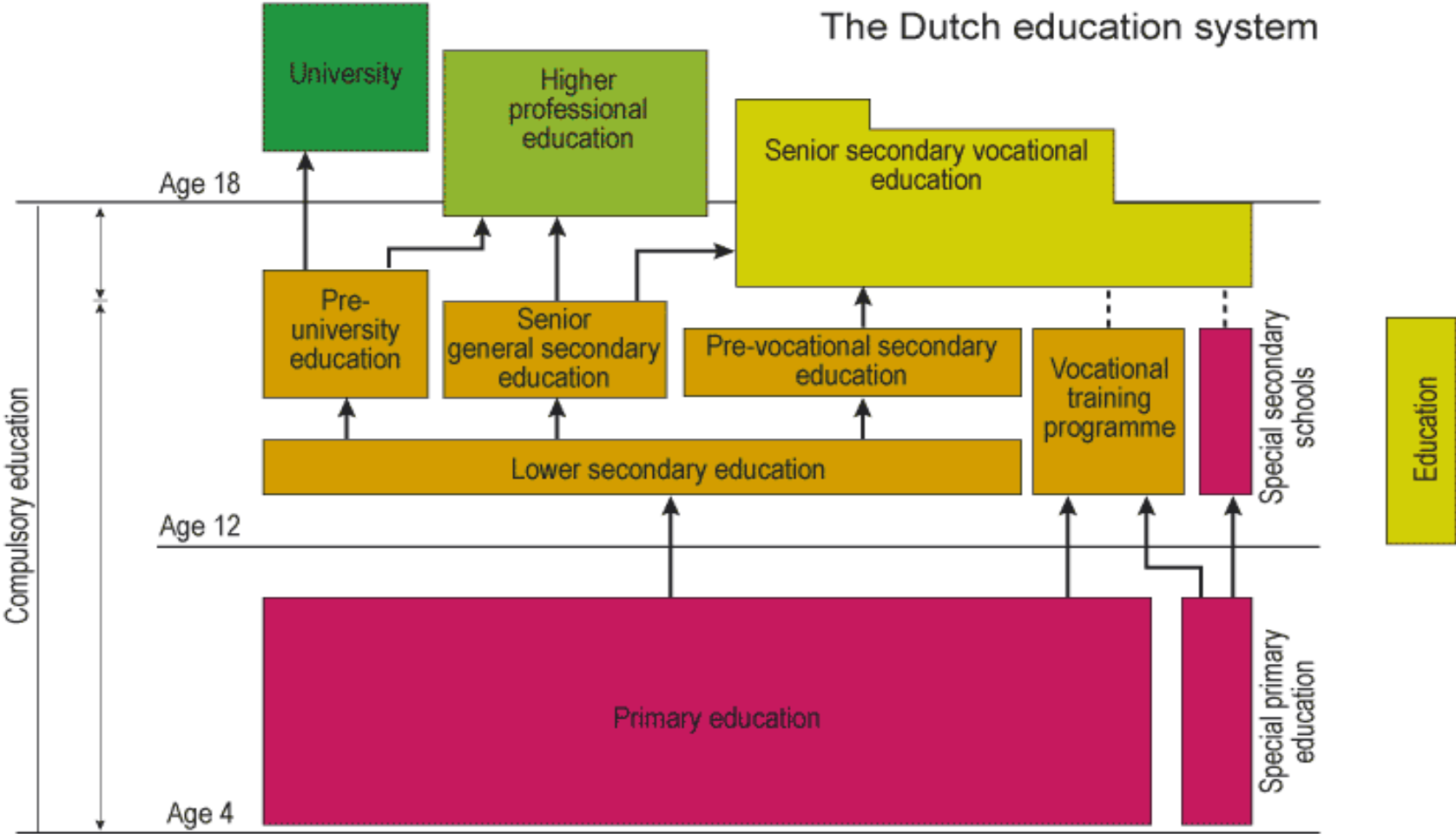
now you know

Cito and Standardized testing

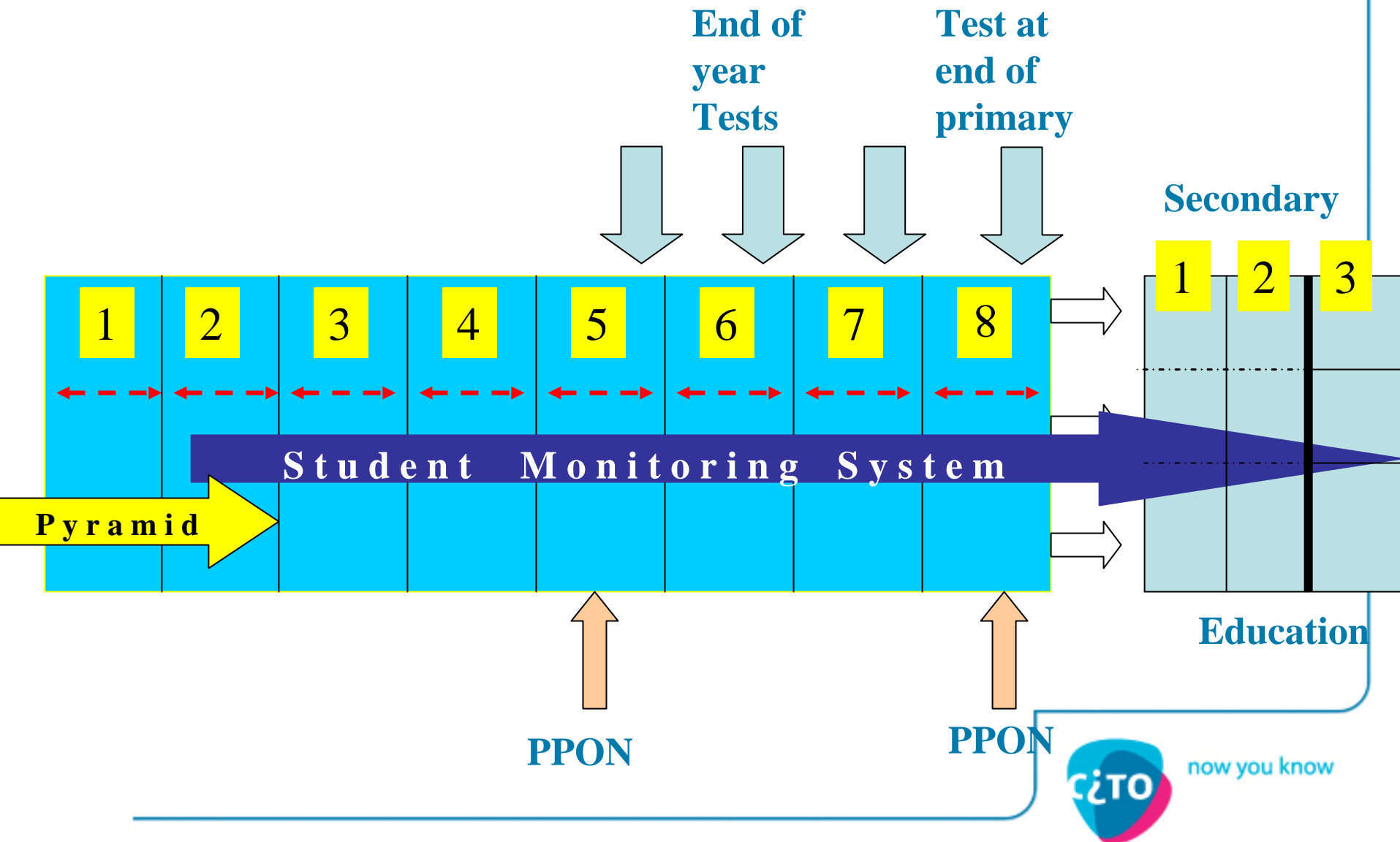
Entreetoets grade 7 ?

- Cito known from 'Cito test'
(School leaving test end of PE, grade 8, age 12)
- => Standardized tests => Cito test
Not only at the end of education but also during.
- Student Monitoring System ('internal evaluation')
- 'External evaluation' 'Cito test' 'Entreetoets'
- Comparability of outcomes SMS Citotest Entree
- Involvement of all stakeholders
Students – Schools – Society (Government)

The Dutch education system



Assessment in schools



Entreetoets Groep 7 (at age 11)

- Purpose: |
 - Provide overview of student profile (paper)
 - Report school achievement (internet)
 - Report ability score (comparability)
- Content:
 - Dutch Language
 - Mathematics
 - Social science
- 16 tasks, 450 items, 13+3 topics, 130.000 students (75% of population)
- Testtaking period: April-September
- Renewed every 5 years (about)

Jaar: 2010
Cito-nr: 12345
School: Citoschool
Plaats: Arnhem

Naam: VAN DEN AKKER, J
Nummer: 123456
Klas: 7A

Printdatum: 11-1-2010

	aantal opgaven	aantal goed	percentiel	V	IV	III	II	I
TAAL	220	148	43			*		
• Schrijven	50	28	31		*			
• Herkennen van de persoonsvorm	25	23	53			*		
• Spelling	60	32	34		*			
- Niet-werkwoorden	30	22	54			*		
- Werkwoorden	30	10	20		*			
• Leestekens								
• Begrijpend luisteren								
• Begrijpend lezen	50	38	43		*			
• Leestempo								
• Woordenschat	60	50	84				*	
REKENEN-WISKUNDE	120	69	28		*			
• Getallen en bewerkingen	45	23	17		*			
• Verhoudingen, breuken en procenten	25	17	46			*		
• Meten, meetkunde, tijd en geld	50	29	31		*			
- Meten en meetkundige vraagstukken	30	19	43		*			
- Tijd en geld	20	10	19		*			
STUDIEVAARDIGHEDEN	80	55	66				*	
• Studieteksten	20	11	39		*			
• Informatiebronnen	20	11	39		*			
• Kaartlezen	20	15	70				*	
• Lezen van schema's, tabellen en grafieken	20	91	91					*
TOTAALSCORES								
• TAAL	220	148	43		*			
• REKENEN-WISKUNDE	120	69	28		*			
• STUDIEVAARDIGHEDEN	80	55	66				*	
TOTAAL ENTREETOETS GROEP 7	420	272	42		*			



zeker weten

Jaar: 2010
Cito-nr: 12345
School: Citoschool
Postbus 1034
6801 MG Arnhem

Aantal leerlingen
2010: 22 van 22
2009: 23 van 23

Gemiddeld aantal goed zonder correctie
270,0
265,1

Landelijk gemiddelde correctie LG

	aantal opgaven	ondergrens	gecorrigeerd aantal goed	bovengrens	-2	-1	0	+1	+2
TAAL	220	137,7	147,3	157,0		X			
• Schrijven	50	30,9	33,5	36,2		X			
• Herkennen van de persoonsvorm	25	20,1	22,0	23,8			X		
• Spelling	60	31,9	35,3	38,7		X			
- Niet-werkwoorden	30	19,1	20,8	22,5		X			
- Werkwoorden	30	12,6	14,6	16,6		X			
• Leestekens									
• Begrijpend luisteren									
• Begrijpend lezen	50	34,6	36,9	39,3			X		
• Leestempo									
• Woordenschat	60	39,1	41,8	44,6			X		
REKENEN-WISKUNDE	120	75,0	81,5	87,9		X			
• Getallen en bewerkingen	45	28,6	31,1	33,6		X			
• Verhoudingen, breuken en procenten	25	15,6	17,1	18,6			X		
• Meten, meetkunde, tijd en geld	50	30,6	33,3	36,0			X		
- Meten en meetkundige vraagstukken	30	18,3	20,0	21,7			X		
- Tijd en geld	20	12,2	13,3	14,5			X		
STUDIEVAARDIGHEDEN	80	43,3	47,6	51,8		X			
• Studieteksten	20	9,1	10,1	11,1		X			
• Informatiebronnen	20	9,1	10,1	11,1			X		
• Kaartlezen	20	11,2	12,4	13,6			X		
• Lezen van schema's, tabellen en grafieken	20	13,5	15,0	16,5				X	
TOTAALSCORES									
• Taal	220	137,7	147,3	157,0		X			
• Rekenen-wiskunde	120	75,0	81,5	87,9			X		
• Studievaardigheden	80	43,3	47,6	51,8			X		
TOTAAL ENTREETOETS GROEP 7	420	256	276,0	296,7		X			



zeker weten

Embedded fieldtest 2009

- Embedded field test in 2009
(avoiding pretest effects, 'shopping' behaviour)
- 9 versions, each 200 'old' 250 'new' items
=> optimal design
- IRT 'One Parameter Logistic Model hammer' to
equate these versions
to the 'Entreetoets 7' by topic for reporting 2009
- Construction of new test
- Prepare Reporting system for 2010



now you know

Design embedded fieldtest

- [E:\irt workshop 2010_frans\DesignEntree7-2009-090128.xls](#)

Main characteristics Fieldtest

- Each item in at least two booklets
- Number of observations per item: 600
- Number of pretest students 2700
- New test constructed based on EF+
- Sample of 5000 students 'old' test

- Was this good enough?
 - Number of observations in pretest
 - Stability of item parameters from pretest
- Work in progress.....

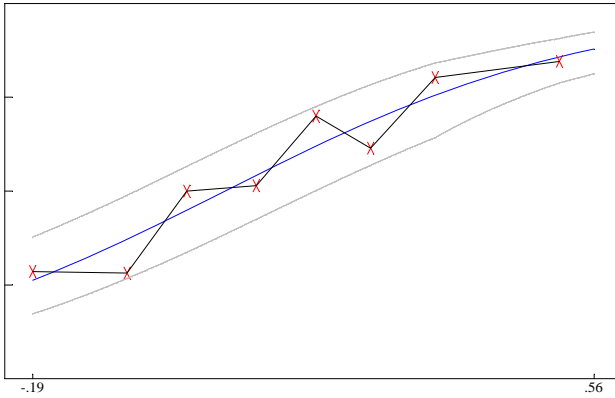


Evaluation

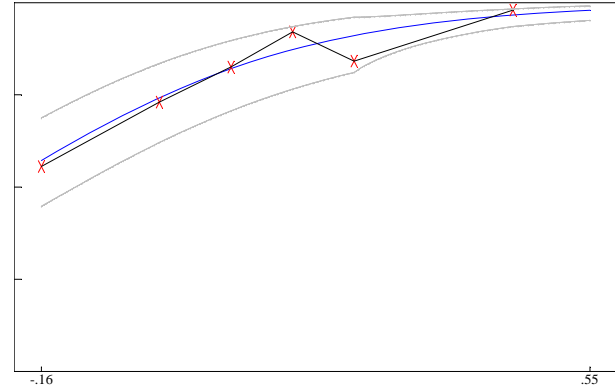
- IRT model Embedded Fieldtest – EF for subscales
- Add test of 2010 test onto fixed scale of EF sample of 5000

Maths: Typical Itemcurves beginning and end of tasks

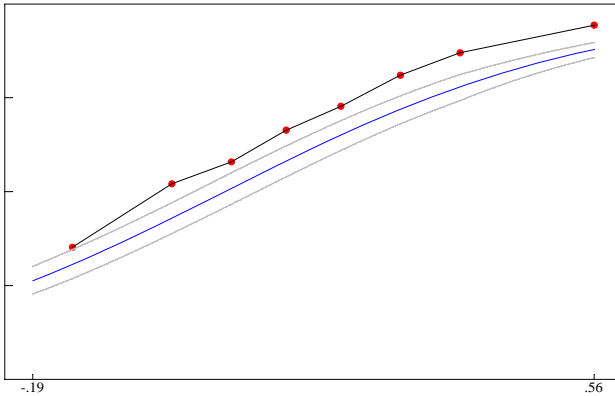
Rel. item #: 222 Abs. item #: 787 Label: 77055 [-1] PTE7



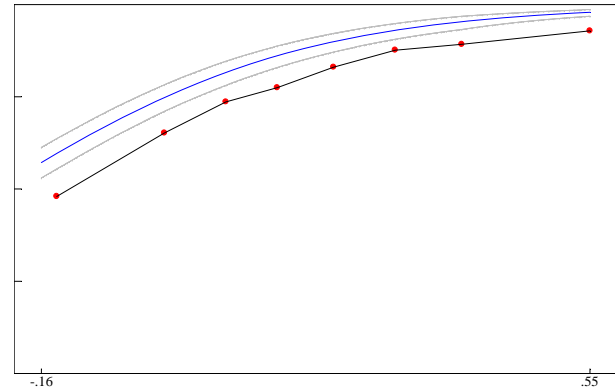
Rel. item #: 247 Abs. item #: 904 Label: 77029 [-1] PTE7



Rel. item #: 222 Abs. item #: 787 Label: 77055 [-1] E7_10



Rel. item #: 247 Abs. item #: 904 Label: 77029 [-1] E7_10



Evaluation

- DIF: Calibration scale - Test 2010 misfit on fixed scale
- Some misfit caused by guessing
- Item sequence seems to matter

First results of evaluation

- Is 600 observations in pretest sufficient?
- Model item sequence?
 - => we do not really want this
 - => consequences for design of pretest
 - => consequences for Adaptive testing
- Advise: Do embedded fieldtest twice?
 - => Once to select items
 - => Once in intended sequence



For further information:
www.cito.nl

frans.kleintjes@cito.nl