Diagnostic testing in Facet

Innovations in computer based testing

Johanna Hofstee
Project manager,
Cito, Institute for Educational Measurement

Nynke de Boer
Manager computer based testing,
CvTE, Dutch Exam Board
Outline

- National exams and tests in the Netherlands
- Facet
- The diagnostic midterm test (DTT)
- Innovations in computer based diagnostic testing
  - Item types
  - Scoring
  - Adaptivity
  - Reporting
- Maintaining standards
Who are we?
Education system in the Netherlands

- Elementary school
- VMBO Pre-vocational
- HAVO
- HBO Bachelor
- MBO Vocational
- WO Bachelor
- VWO

Age:
- 22
- 20
- 18
- 16
- 14
- 12
- 10
- 8
- 6
- 4
National exams and tests

- Dutch as a second language exam
- Literacy and Numeracy exam
- Final exam
- Numeracy exam
- DTT: Diagnostic midTerm Test
- Final test

Age:
- .... 22
- .... 20
- .... 18
- .... 16
- .... 14
- .... 12
- .... 10
- .... 8
- .... 6
- .... 4

CBT
Paper
CBT deliveries 2015-2016

1.200.000 deliveries and growing to 2-3M
3000 locations, growing to 5000-8000
Developing Facet

For delivering 2-3 million CBT per year we needed a new system

- Stable, well secured, able to allow large numbers
- Open source and multi platform
- Proper item display – pixel perfect!
- Minimal effort for schools
- Online and offline
- Fit within the governmental architecture, using open standards where possible
- Possible to link to existing governmental systems
- No exceeding of costs!

Decision was made in 2007 to create our own system
## Development Facet

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Handmatig inplannen</td>
<td>X</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offline afname</td>
<td>X</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examenbestanden inlezen/previewen</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normale (QTI) vraagtypes</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verklanking</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correctiemodule</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Online afname</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oefenomgeving</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptief toetsen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innovatieve taalitems (DTT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Innovatieve rekenitems (DTT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Maximale opschaling (cloud)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Toetsen over dagen heen (PO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Uitgebreide toegankelijkheid (alle)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Bijstellen datamodel (alle)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spraak als antwoord (NT2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Gesegmenteerd beoordelen (NT2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
Interesting challenges…

- Multi platform, multi-device
- Working in the cloud
- Keeping up with new privacy laws
- Making Facet available for third parties (as in an Open Source community?)
• QTI as exchanging standard for content
• DEP – Dutch Exam Profile
• Looking into other IMS standards, like Caliper, APIP and LTI
• Support IMS in Europe together with partners like Cito, OAT, Trifork and Surf.
• Current initiatives:
  • IMS Europe board
  • QTI-board EBA
  • IMS events in Europe
    • 16-18 november in the Netherlands
National exams and tests

- Dutch as a second language exam
- Literacy and Numeracy exam
- Final exam
- Numeracy exam
- **DTT: Diagnostic midTerm Test**
- Final test
• **Construction of assessments**
  • Exams and tests ~ Dutch Exam Board
  • Range of other assessments

• **Psychometrics**
  • Calibration – equating – standard setting – optimal test design

• **Technology**
  • CBT-platform ‘Questify’: authoring, item banking, construction and delivery of computer based tests
Why the DTT? (diagnostic midterm test)

- Government policy (2011): part of a plan of action to improve Dutch secondary education

- Improve students’ results on core subjects: Dutch, English and mathematics
• Diagnosis on the students’ strengths and weaknesses in Dutch, English and mathematics

• Halfway to their final exams

• Information to improve learning and teaching

• Innovative, computer based test, delivered in Facet
How?

- Under direction of the CvTE
  - Content development by Cito
  - Development Facet by DUO
  - Development adaptive module by Cito

- Initial plan: mandatory test for all students (ready 2015)

- Now: pilot to deliver a full prototype, in cooperation with schools (ready 2017)
Provide a reliable diagnosis on all aspects of the student’s skill, in tests of a reasonable length.
Basic ingredients (1): skill models

- Which (sub)skills are necessary to yield a performance at the required level?
- For example: English writing

<table>
<thead>
<tr>
<th>Social proficiency</th>
<th>1.1. Can use the appropriate tone of voice/register</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.2. Can use the appropriate conventions for the text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structuring</th>
<th>2.1. Can structure a text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.2. Can use appropriate structuring words (referers)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phrasing</th>
<th>3.1. Can use appropriate words/word combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.2. Can vary vocabulary where appropriate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcribing</th>
<th>4.1. Can use correct word order and grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.2. Can use correct spelling and punctuation</td>
</tr>
</tbody>
</table>
Basic ingredients (2): report model

Diagnosis in terms of:
- under level
- on level
- above level

### DIAGNOSE van
**Schrijfvaardigheid Engels**

<table>
<thead>
<tr>
<th>DIAGNOSTISCH PROFIEL</th>
<th>VERDIEPENDE DIAGNOSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afstemming op publiek en doel</td>
<td>Gebruik basisconventies</td>
</tr>
<tr>
<td>Coherentie</td>
<td>Gebruik voegwoorden en verwijzwoorden</td>
</tr>
<tr>
<td>Woordenschat en woordgebruik</td>
<td>Gebruik passende woorden en woordcombinaties</td>
</tr>
<tr>
<td>Spelling, interpunctie en grammatica</td>
<td>Passende spelling en interpunctie</td>
</tr>
</tbody>
</table>
Basic ingredients (3): CBT-chain

- QuestifyBuilder
- DEP-QTI Testpackage
- Testdesign
- Facet Administration
- Facet Testplayer
- Facet Scoring
- Facet Dataserver
- Resultdata
- Data-analysis tools
- Facet Reporting
What we needed (1): items and scoring

- Items that directly address the student’s skill
- Items that yield as much information as possible
- All responses automatically scored, constructed-responses included!
- Items that generate data in terms of the skill models
New features 1: new item types

- construction and delivery of new item types (~QTI)
New features 2: encoding responses

Item authoring includes ‘encoding’ of all possible responses: define the diagnostic value

Encoding step 1:

- (in)dependent interactions?

![Image of score editor with options for score method, field name, operator, and value.

Score method: Dichotomous

<table>
<thead>
<tr>
<th>Field name</th>
<th>Operator</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>=</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>=</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>=</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>=</td>
<td>D</td>
</tr>
</tbody>
</table>
New features 3: encoding responses

Encoding step 2:

- relate all possible responses to one or more (sub)aspects of the skill model

### Encoding editor

<table>
<thead>
<tr>
<th>Name</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ENG-1.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ENG-1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENG-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Social proficiency

1.1. Can use the appropriate tone of voice/register
1.2. Can use the appropriate conventions for the text

#### Structuring

2.1. Can structure a text
2.2. Can use appropriate structuring words (referers)

#### Phrasing

3.1. Can use appropriate words/word combinations
3.2. Can vary vocabulary where appropriate

#### Transcribing

4.1. Can use correct word order and grammar
4.2. Can use correct spelling and punctuation
Generating diagnostic data
- more 'observations' per item
- related to skill model

<= data used in diagnostic test
<= 'normal' data is not relevant
New features 4: custom interactions

• interactive mathematics
• interactions created in external applications
New features 5: all items automatically scored

- Scoring of mathematical expressions bij a computer algebra system (CAS, Maxima)
- Selection of scoring types in authoring tool
New features 6: scoring by CAS

- CAS-scoring types
- equal strict
New features 7: scoring by CAS

- CAS scoring types
- equivalent
CBT-innovations: items and scoring

- New item types
  - QTI-based
  - Constructed-responses
  - Custom interactions

- Automatic scoring of all responses

- Generating ‘diagnostic’ data ~ encoded responses
What we needed (2): delivery and reporting

- Testdelivery yields a classification (under, on or above level) of students on many sub-skills

- Time-reduction by adaptive testdelivery!

- Item-interaction parameters based on pretestdata:

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Item</th>
<th>p1</th>
<th>p2</th>
<th>p3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCEPTRESPONSE_ENG-1-1</td>
<td>dtt-en-vm-0395-SC</td>
<td>0.18987350</td>
<td>0.29605270</td>
<td>0.42148760</td>
</tr>
<tr>
<td>CONCEPTRESPONSE_ENG-1-2</td>
<td>dtt-en-vm-0423-SC</td>
<td>0.50632911</td>
<td>0.70394730</td>
<td>0.87603294</td>
</tr>
<tr>
<td>CONCEPTRESPONSE2_ENG-1-2</td>
<td>dtt-en-vm-0423-SC</td>
<td>0.10126590</td>
<td>0.26315793</td>
<td>0.44628100</td>
</tr>
</tbody>
</table>

- Online reporting of usable diagnoses
New features 8: adaptive test construction

An adaptive test contains:

- **Items**
- **Adaptive driver:**
  - item parameters
  - test design (item blocks)
  - stop criterium
  - report criterium
- **Adaptive module, with psychometric core**
New features 8: Testdesign with blocks

- Itemblocks designed classify students
  - as reliable as possible
  - on - under - above level
  - on each (sub) aspect

![Diagram of itemblocks classification](image)
New features 10: adaptive course

- Diagnosis in two steps

Section 1: diagnoses of main attributes

Section 2: diagnoses of sub-attributes

If indicated and time
New features 11: adaptive test delivery

1. Present first item
2. Answer item
3. Score item (CAS)
4. Present item
5. End test

Loop

First item
Response
Next item

Adaptive Engine

Update report after each response

Adaptive report based on item statistics and student response
New features 12: result data

Resultdata:

- in terms of the skill model
- probability of the classification >= 0.8

<table>
<thead>
<tr>
<th></th>
<th>P1 (under)</th>
<th>P2 (on)</th>
<th>P. (above)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG-1</td>
<td>8.59523e-55</td>
<td>2.57987e-07</td>
<td>1</td>
</tr>
<tr>
<td>ENG-1-1</td>
<td>5.39044e-24</td>
<td>0.0013884</td>
<td>0.998612</td>
</tr>
<tr>
<td>ENG-1-2</td>
<td>4.11786e-13</td>
<td>0.0292907</td>
<td>0.970709</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...
### New feature 12: diagnostic reporting

**School:** School XYZ  
**Klas:** Klas 3B  
**Leerweg:** havo  
**Vak:** Engels

#### Diagnostic van Schrijfvaardigheid Engels

<table>
<thead>
<tr>
<th>Klas</th>
<th>Leerling</th>
<th>Score 1</th>
<th>Score 2</th>
<th>Score 3</th>
<th>Score 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3B</td>
<td>Naam 0156</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0160</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0170</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0172</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0166</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0154</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0151</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0153</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0168</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0169</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0171</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0162</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0155</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0159</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0164</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0163</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0157</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0152</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0161</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0165</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0158</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0174</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0175</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>Naam 0173</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Diagnosticch Profiel

1. Afstemming op publiek en doel  
2. Coherentie  
3. Woordenschat en woordgebruik  
4. Grammatica, spelling en interpunctie
CBT-innovations: delivery en reporting

- Adaptive module, integrated in Facet delivery
- Reporting module DTT
Conclusion: diagnostic testing asks for …

• efficiency in all aspects of test construction and delivery!

• dig deeper, get more information

• combination of innovations in CBT-platform
Invitation

Description
International symposium on diagnostic testing in education

Date and time
Start: 8th of February 2017, 12.00 PM
End: 10th of February 2017, 13.00 PM

Place
Mitland Hotel, Utrecht, the Netherlands

More information?

dtt@hetcvte.nl

joke.hofstee@cito.nl

n.deboer@hetcvte.nl