UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Adaptive Test Development Using Concerto Platform

Dr Luning Sun

Psychometrics Module, Lecture 4

Social Sciences Research Methods Centre | SSRMC

Introduction to CAT

Some materials and examples come from previous workshops run by:

Michal Kosinski (Stanford University)

David Stillwell (University of Cambridge)

Chris Gibbons (Harvard University)

Computerised Adaptive Testing

- Standard test is likely to contain questions that are too easy or too difficult
  - Classical Test Theory vs Item Response Theory
- Adaptively adjusting the level of the test to individual participant:
  - Increases the accuracy
  - Saves time / money
  - Prevents frustration

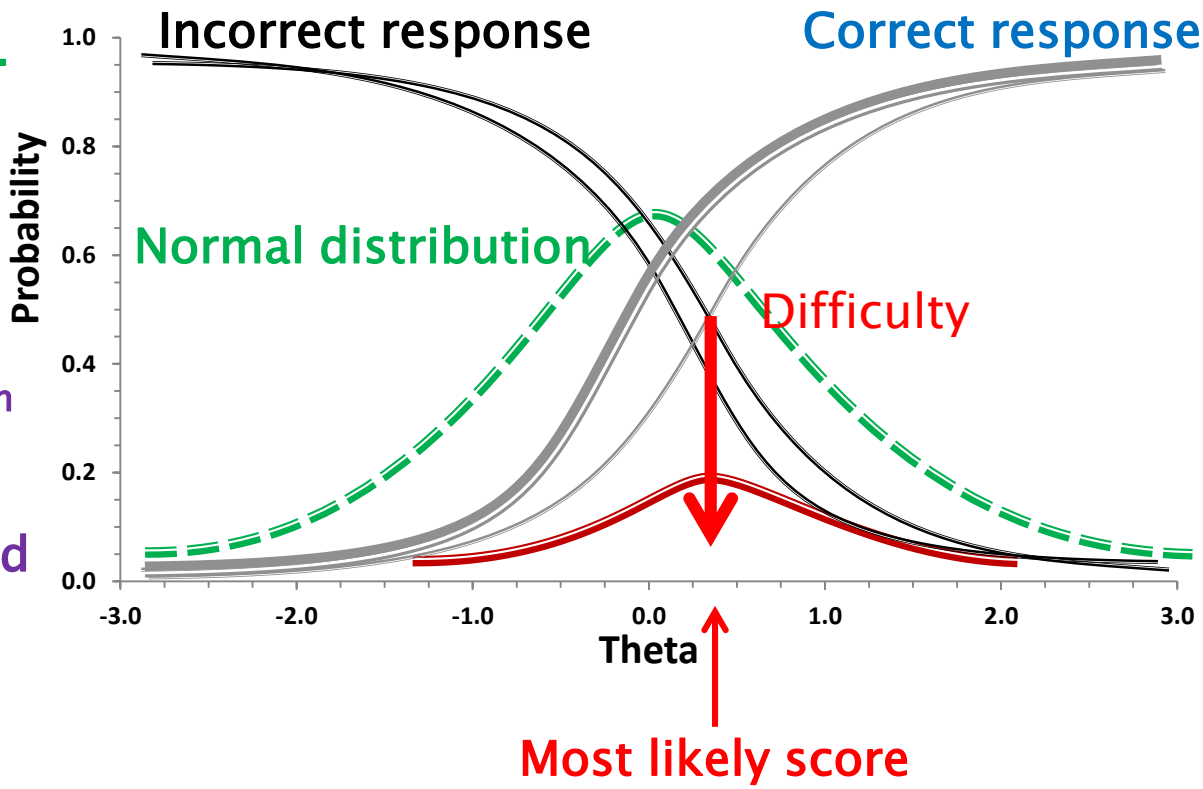
Elements of CAT

- IRT model
- Item bank and calibration
- Starting point
- Item selection algorithm (CAT algorithm)
- Scoring on-the-fly method
- Termination rules
- Item bank protection / overexposure
- Content Balancing

Example of CAT

Start the test:

1. Ask first question, e.g. of medium difficulty
2. Correct!
3. Score it
4. Select next item with a difficulty around the most likely score (or with the max information)
5. And so on…. Until the stopping rule is reached



Classic approaches to item selection

- Maximum Fisher information (MFI)
  - Obtain a current ability estimate
  - Select next item that maximises information around the current ability estimate
- Urry's method (bOpt; in 1PL equals MFI)
  - Obtain a current ability estimate
  - Select next item with a difficulty closest to the current one
- Other methods:
  - Minimum expected posterior variance (MEPV)
  - Maximum likelihood weighted information (MLWI)
  - Maximum posterior weighted information (MPWI)
  - Maximum expected information (MEI)

Examples of item overexposure prevention

- Randomesque approach (Kingsbury & Zara, 1989)
  - Select >1 next best item
  - Randomly choose from this set
- Embargo on overexposed items
- Location / Name / IP address rules

- Large item bank
- Regularly updated item bank

Kingsbury, G. G., and Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. Applied Measurement in Education, 2, 359-375.

Content Balancing

- Ascertain that all subgroups of items are used equally
- Example:
  - Arithmetic, Algebra and Geometry in a math test
  - Different domains in an intelligence test
  - Emotion recognition test
- Multidimentional CAT

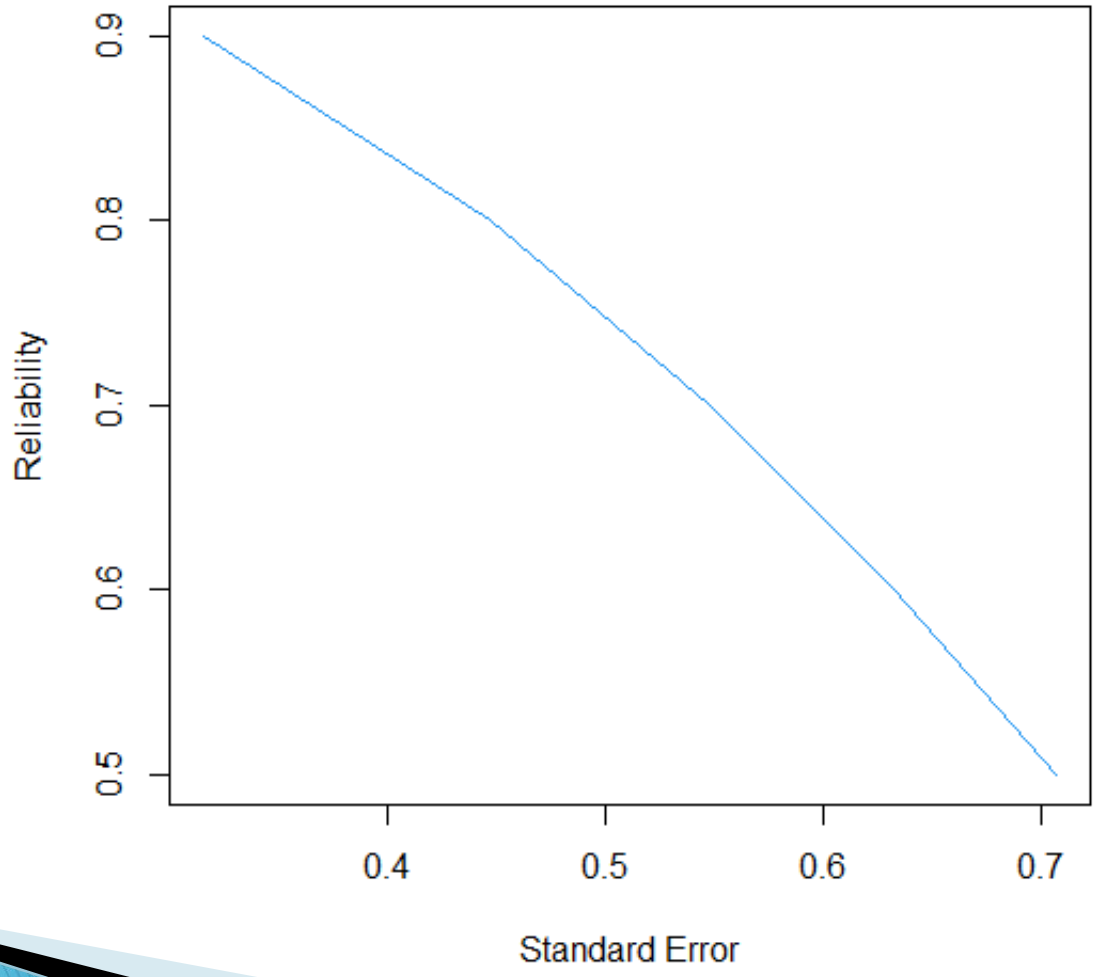
Stopping rules

- Test length (*e.g.*., 20 items, 15 items)
- Test time (5 minutes)
- Reliability of theta estimate (standard error)

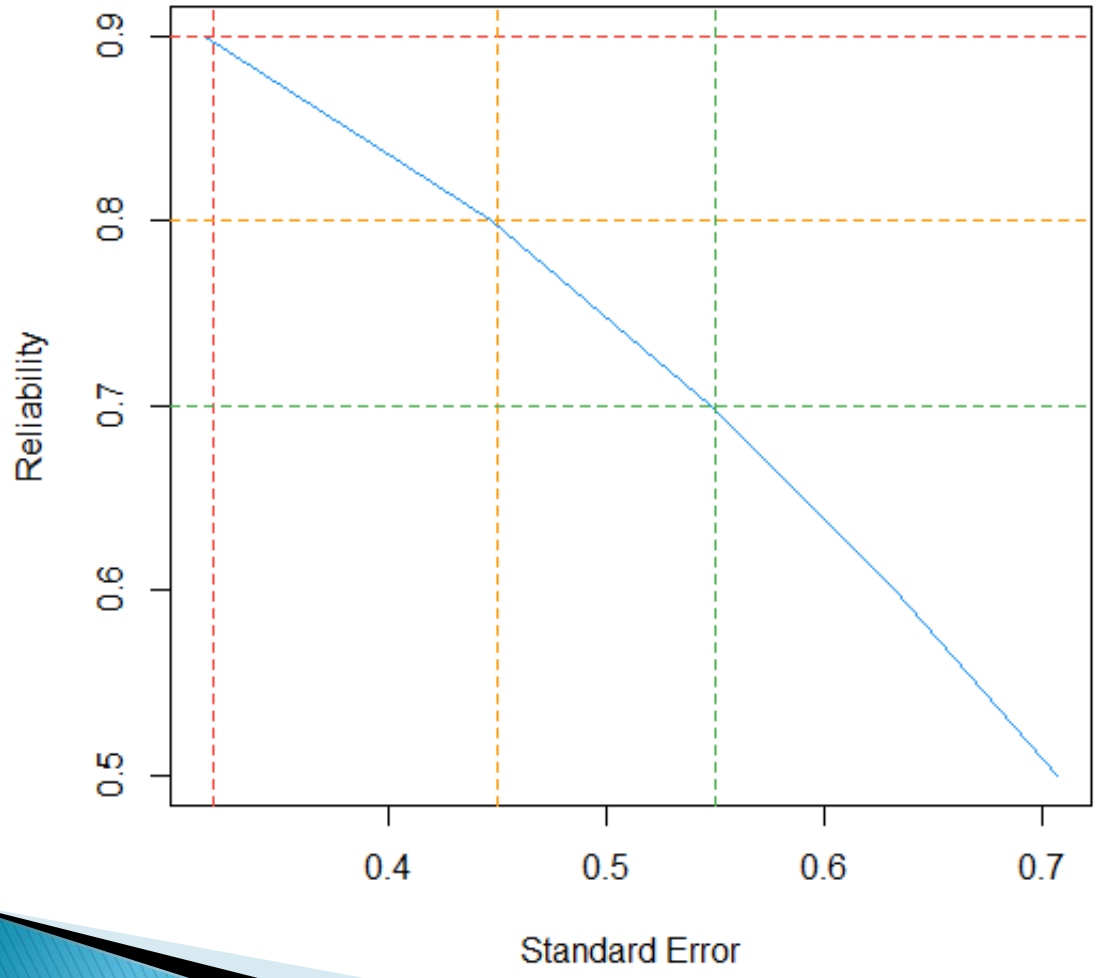
- Other, clever stuff

Reliability and Standard Error

$$reliability = 1 - SE^2$$



Reliability and Standard Error



Alpha(0.90) = SE(0.32)

Alpha(0.80) = SE(0.45)

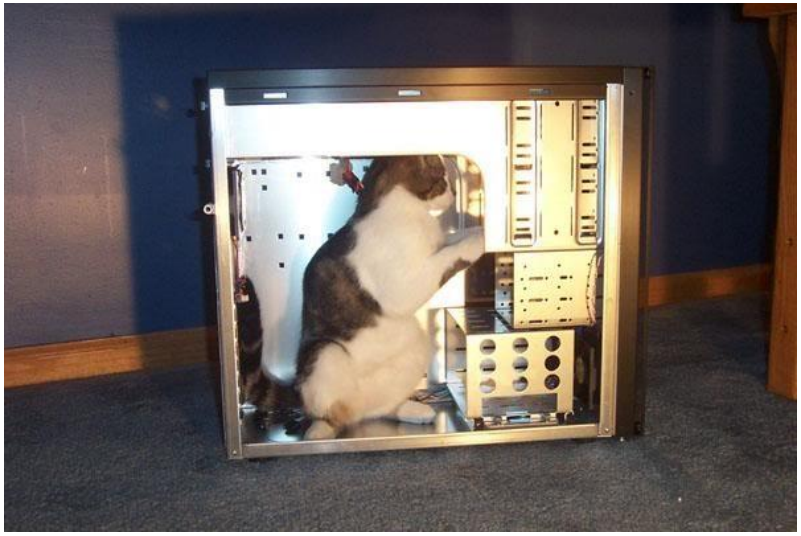
Alpha(0.70) = SE(0.55)

CAT Procedure

1. The pool of available items is searched for the optimal item, based on the current estimate of the examinee's ability
2. The chosen item is presented to the examinee, who then answers it correctly or incorrectly
3. The ability estimate is updated, based upon all prior answers
4. Steps 1–3 are repeated until a termination criterion is met

CAT Qualities

- Efficiency – how many items do I need to ask before I get to a certain level of precision
- Precision – How precise can my measurement be



CAT Summary

- What do we need for CAT –

Item information (questions, scoring keys)
Item parameters
Item selection method
Scoring algorithm
Stopping rule
Others ……

CAT in R

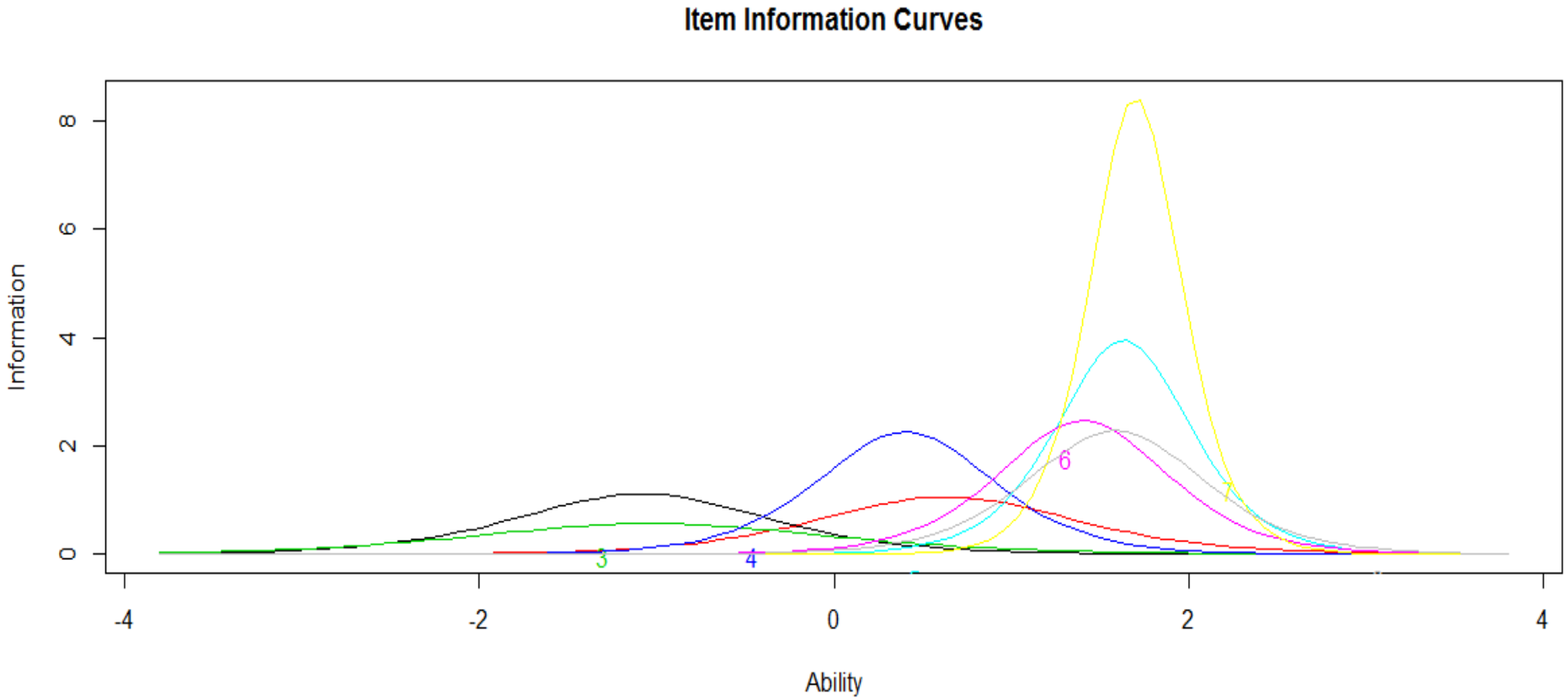
catR package

Example

- Women’s Mobility
  - Item 1Go to any part of the village/town/city.
  - Item 2Go outside the village/town/city.
  - Item 3Talk to a man you do not know.
  - Item 4Go to a cinema/cultural show.
  - Item 5Go shopping.
  - Item 6Go to a cooperative/mothers' club/other club.
  - Item 7Attend a political meeting.
  - Item 8Go to a health centre/hospital.

ltm

```
library(ltm)
my2pl<-ltm(Mobility~z1)
plot(my2pl,type="IIC")
```


Item Information Curves
Information
0
2
4
6
8
-4
-2
0
2
4
Ability
3
4
6
7

catR

```
require(catR)
c<-coef(my2pl)
itemBank <- cbind(c[,2], c[,1], 0, 1)
```

catR

Choose the item to start with:

- **max info around average?**

```
plot(my2pl, type = "IIC")
plot(my2pl, type = "IIC", items=4)
```

- **Random one?**

catR

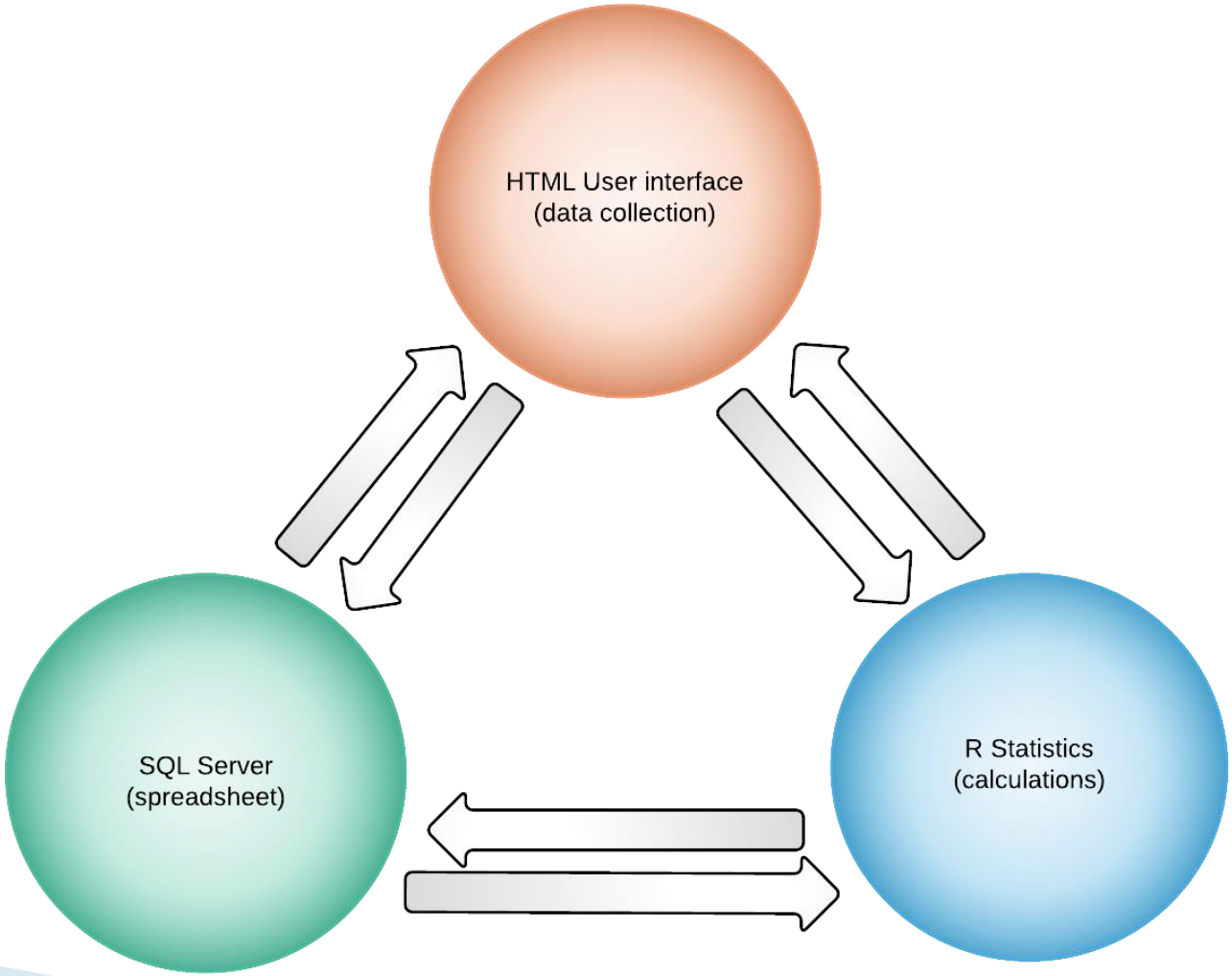
```
items_administered<-c(4)
responses<-c(1)

it<-itemBank[items_administered, 1:4,drop=F]
theta<-thetaEst(it, responses)
sem<-semTheta(theta,it)

q<-nextItem(itemBank, theta=theta,out=items_administered)
q$item
```

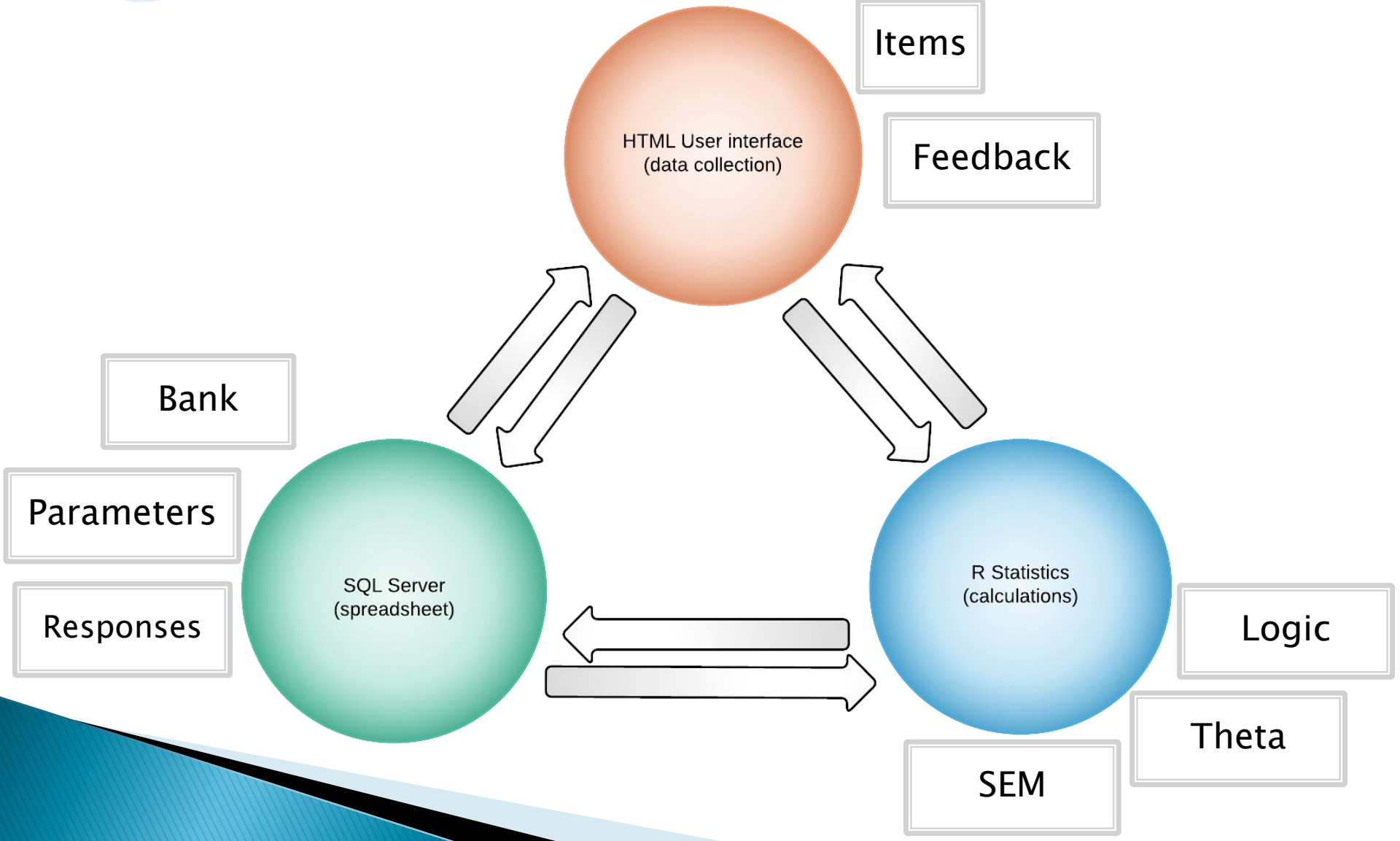
CAT in Concerto

CONCERTO


HTML User interface
(data collection)
SQL Server
(spreadsheet)
R Statistics
(calculations)

CONCERTO

Introduction



Items

Feedback

Bank

Parameters

Responses

Logic

Theta

SEM

CONCERTO

- Concerto hosting website
  - hosting.concertoplatform.com
- Sign up and log in
- Create your own server
- Start your Concerto experience

CONCERTO

V5.0.beta.3.5

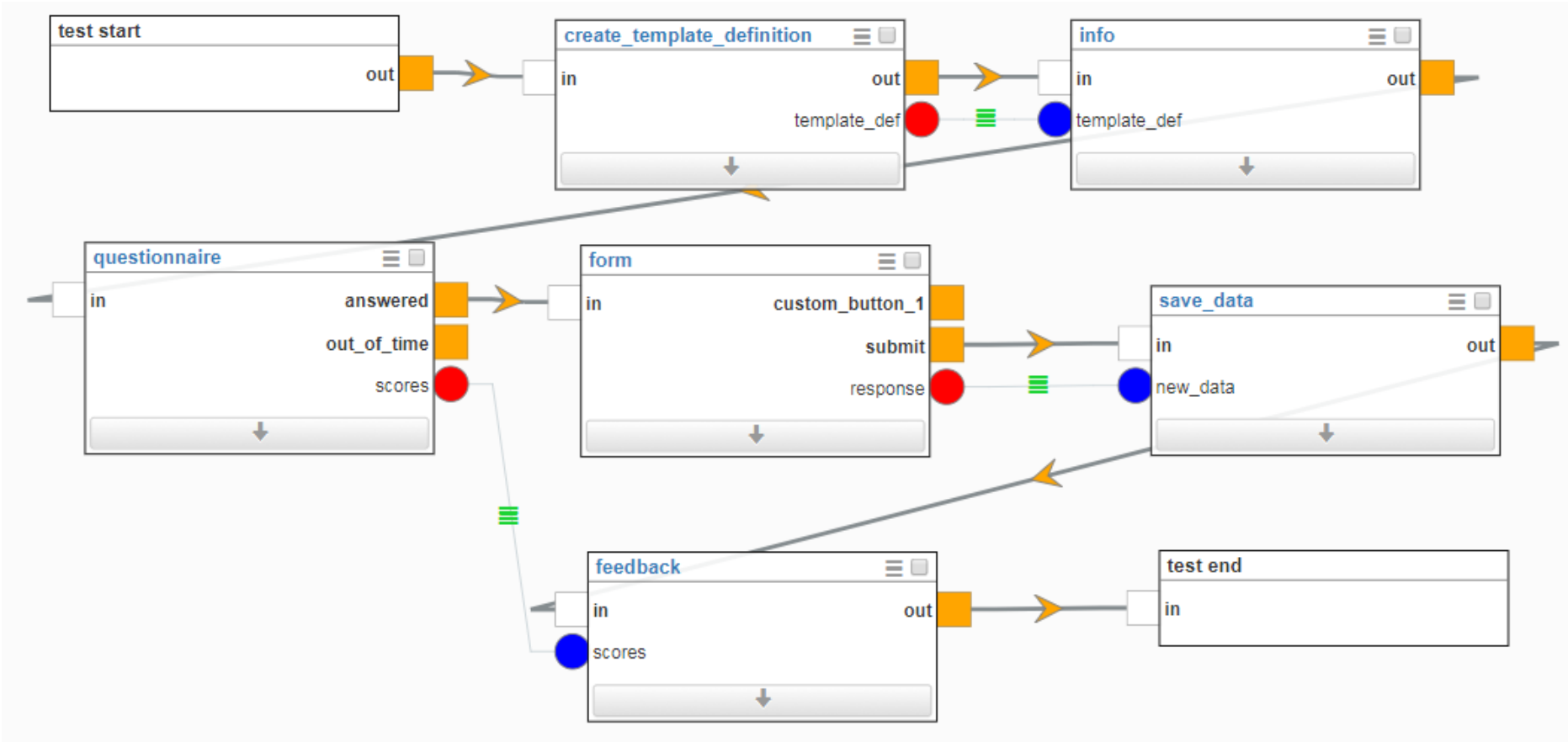
- Name
- URL
- Node:
  - info
  - questionnaire
  - CAT
  - form (save_data)
  - feedback

Practical

- Basic questionnaire

- CES-D scale (The Center for Epidemiologic Studies Depression Scale; Radloff, 1977)
  - 20 items
  - 4 response options
  - Score above 16 indicates depression
- https://concertotest.com/luning/SSRMC/test/cesd

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, *1*(3), 385-401.


test start
out
create_template_definition
in
out
template_def
info
in
out
template_def
questionnaire
in
answered
out_of_time
scores
form
in
custom_button_1
submit
response
save_data
in
out
new_data
feedback
in
out
scores
test end
in

Practical

- CAT – dichotomous

- Women's Mobility
  - 8 items in the item bank
  - Item selection: MFI
  - Scoring: BM
  - Stopping: 3 items
  - Randomesque: 1
  - Content balancing: no
  - Feedback:
    - score$score<-round(score$theta*15+100,0)

FaceIQ

- faceiq.icar-project.com
  - Adaptive face detection test
  - Adaptive emotion recognition test
  - Adaptive abstract reasoning test
  - And more ……

Thank you!

- Any questions?