# Advanced technologies for e-Testing

Maomi Ueno
University of Electro-Communications
Japan

# 1. e-Testing as an advanced technology

- Is e-Learning an advanced technology?
- In other words, does e-learning have unique key technologies which only experts can develop?
- Maybe no.. Almost e-learning system can be developed by ordinal programmer.
- Contrary to e-learning, E-Testing has unique key technologies which only a few experts can develop.

# CBT?

- There have been CBT(Computer Based Testing).
- Is e-Testing different from CBT?

- CBT came from

  Psychology × Statistics

# E-Testing

Psychology × Statistics × Computer Science

Additional problems; How to develop a efficient algorithm to reduce the computational costs.

# 2. Key technologies of  e-Tesing

# 2.1.Item response theory (ex. Lord and Novick 64)

Item response theory (IRT) is the most advanced test theory which explains the item characteristics.

The probability of correct answers of 2-parameter logistic model is

$$p(u_{ij} = 1 \mid \theta_j) = \frac{1}{1 + \exp\{-1.7 a_i (\theta_j - b_i)\}}$$

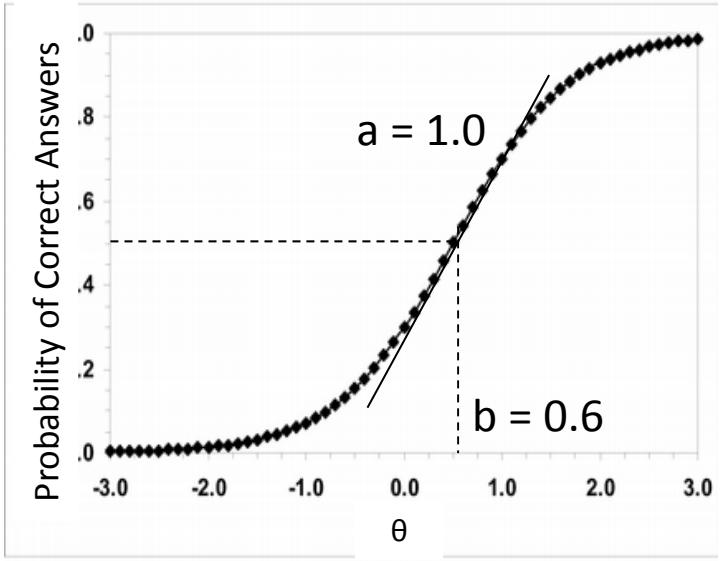
where

- $\vartheta_j$ represents the *j-th* examinee ability
- $a_i$ represents discrimination parameter of the *i* –th item
- $b_i$ represents difficulty parameter of the *i* –th item

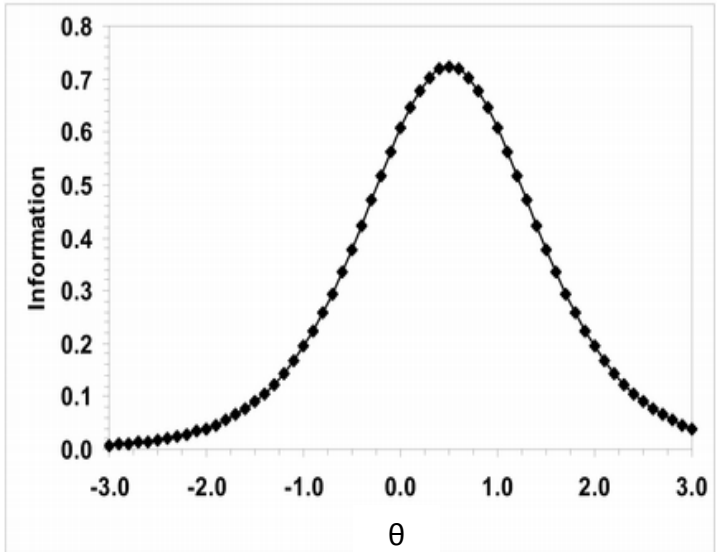
Item information function is

$$I(\theta) = a_i^2 \, p_i(\theta)(1 - p_i(\theta))$$

Item characteristic curve




Item information function



# Advantages

- It makes possible to access examinees who took different test items, on the same scale.

↓

It is very important for high stake tests.

It is available in e-Learning.

# 2.2. Text Mining

- Latent Dirichlet Allocation (Blei et al, 2003)

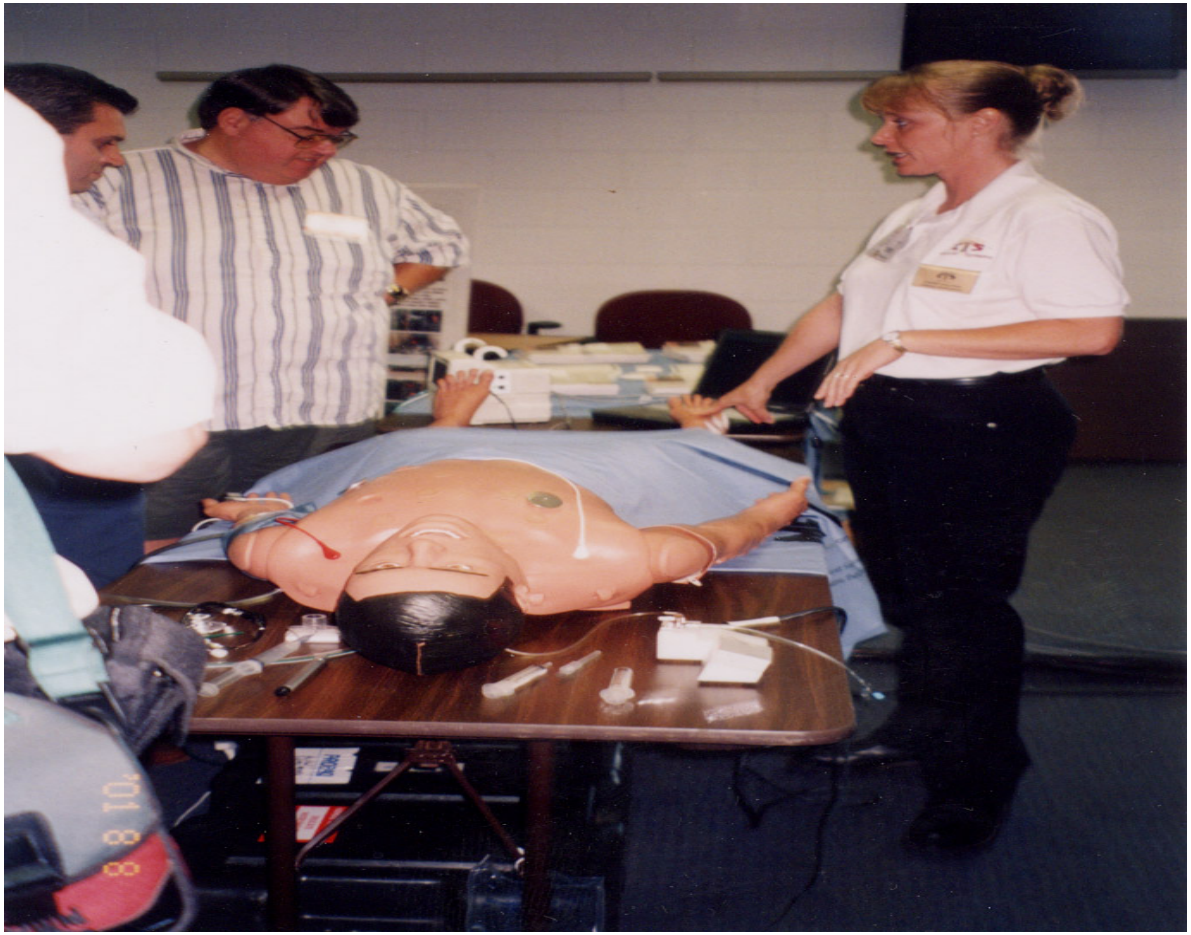
↓

- Automatic construction of item data base
- Automatic Grouping of items
- Automatic Scoring of essay items

# 2.3. User Interface

- Multimedia
- for handicapped examinees :Ex Braille presentation
- Robot

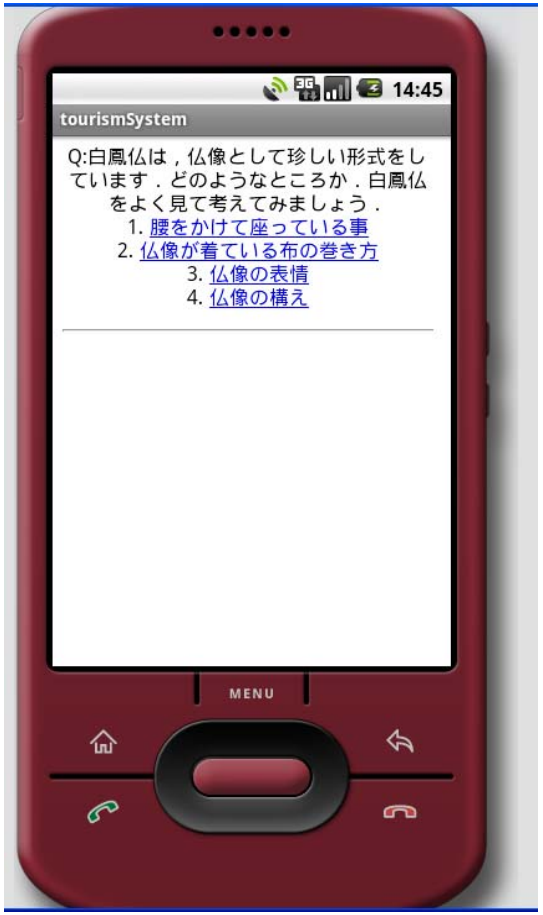
# Medicine Patient Robot



Input the name of a disease and the degree as parameters for Patient Robot

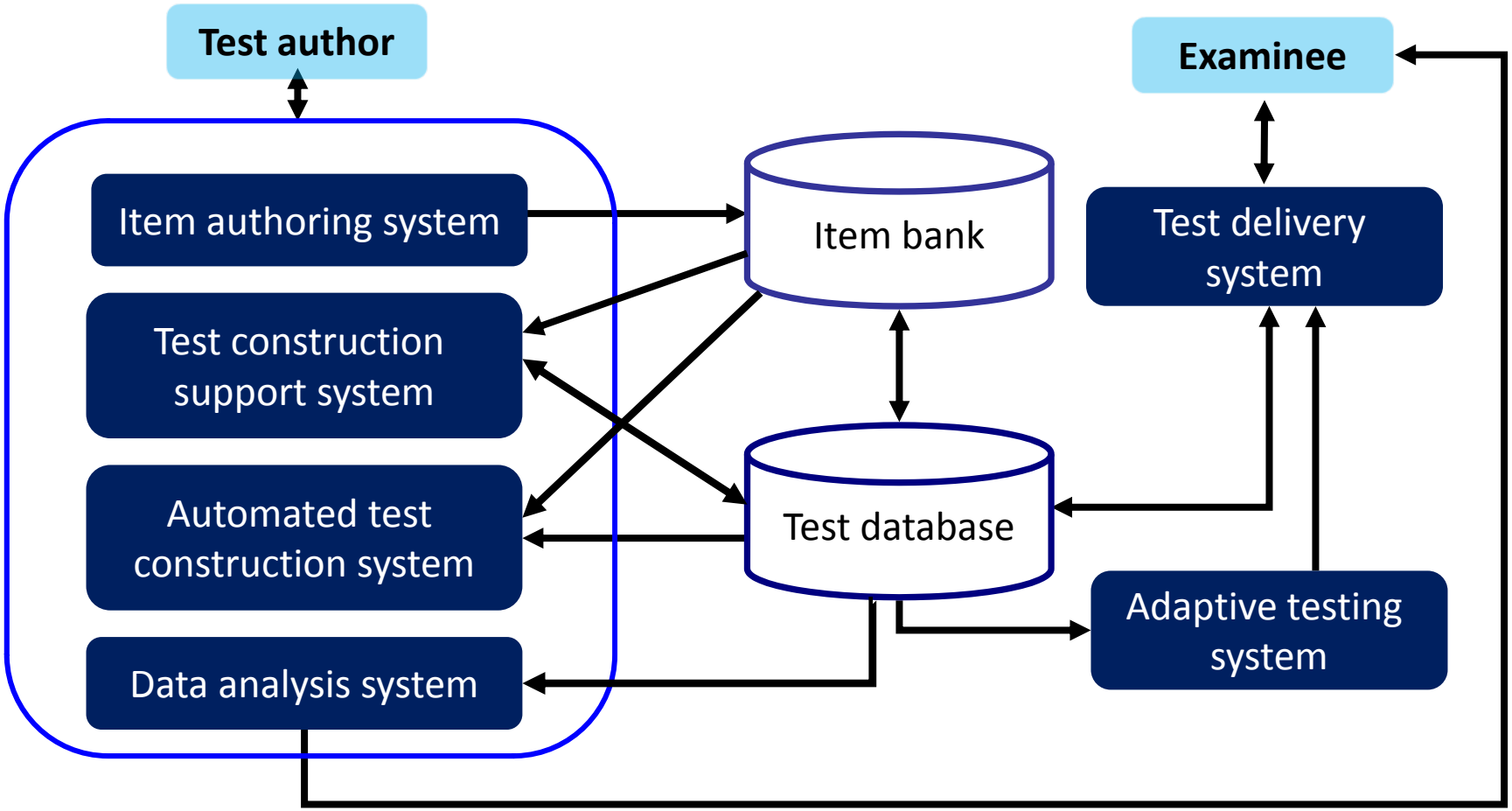
# Mobile Adaptive Testing


Navigation




Adaptive Testing

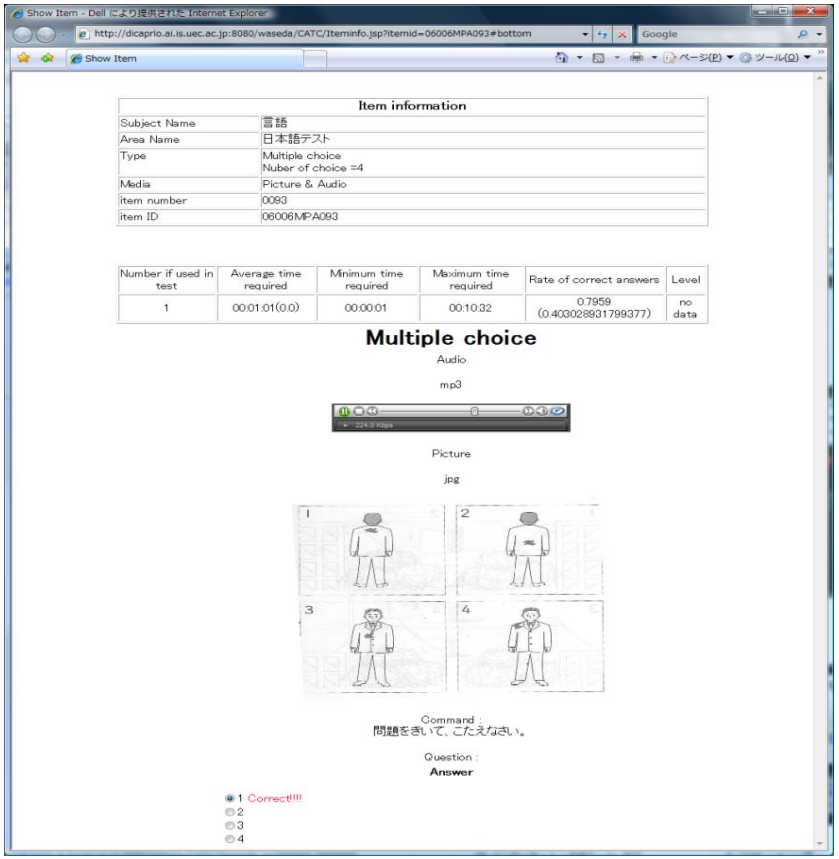
# 3. Outline of e-testing system



# Item authoring system


- Multimedia item
  - Text
  - Image
  - Video
  - Audio
  - Mixing media (for example, image and audio).
- By the e-testing, we can measure the examinee's abilities that have never been measured by the paper-based test.

# Item bank

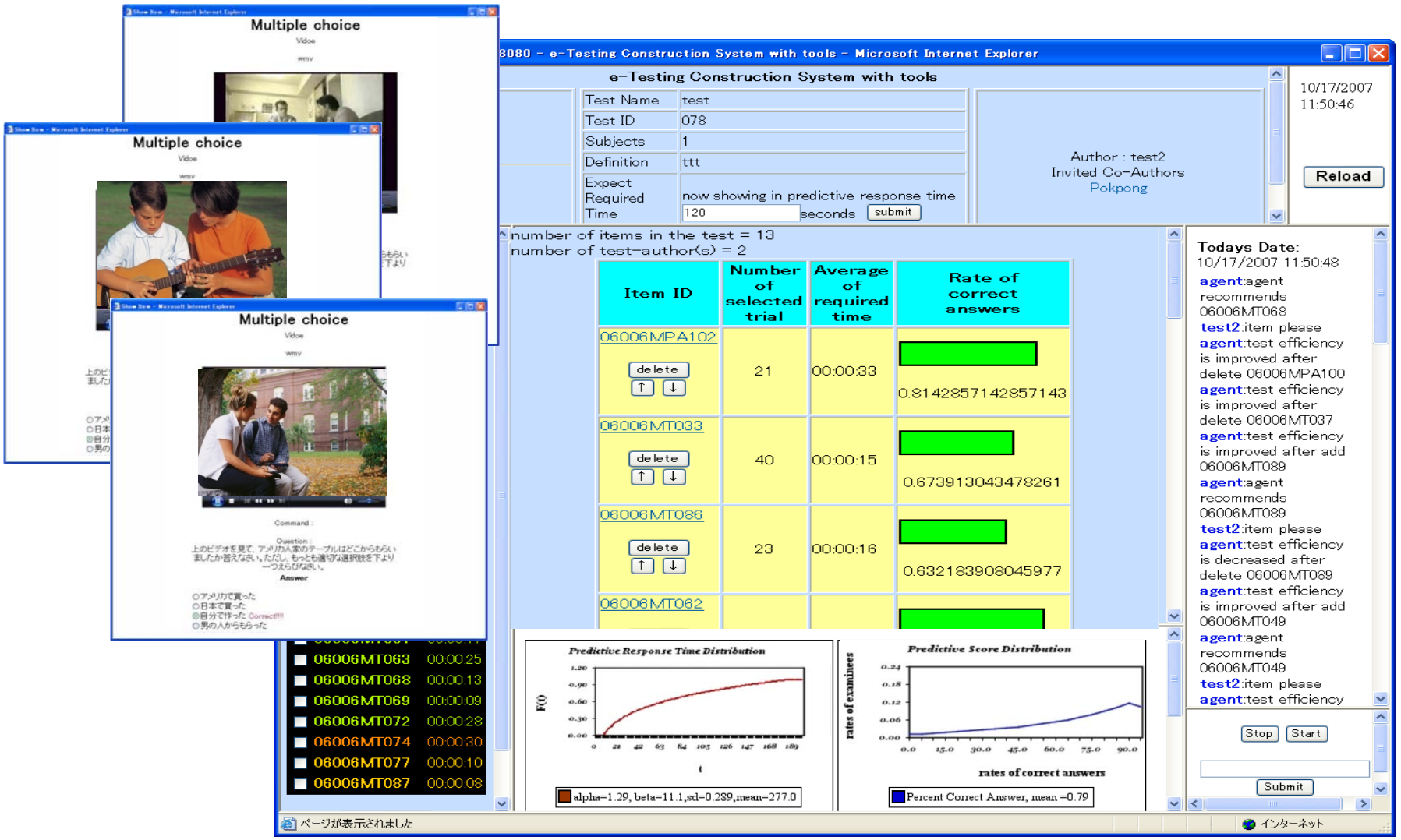



- Information data
  - Item ID
  - Content
  - Media file
- Statistic data
  - Rate of correct answers and standard deviation
  - Average response-time and standard deviation
  - Item response theory parameters (a, b)

# Test construction support system

- In actual test constructions, predictions of a test score distribution and response-time distribution have become important.
- This system provides the prediction tool using the on-line updated item statistics data from the item bank.
- The prediction tool interactively visualizes the score distribution and response-time distribution of the constructed test to support decision making in order to improve the test reliability.
- The system supports collaborative test construction by several test-authors who live in distance places.

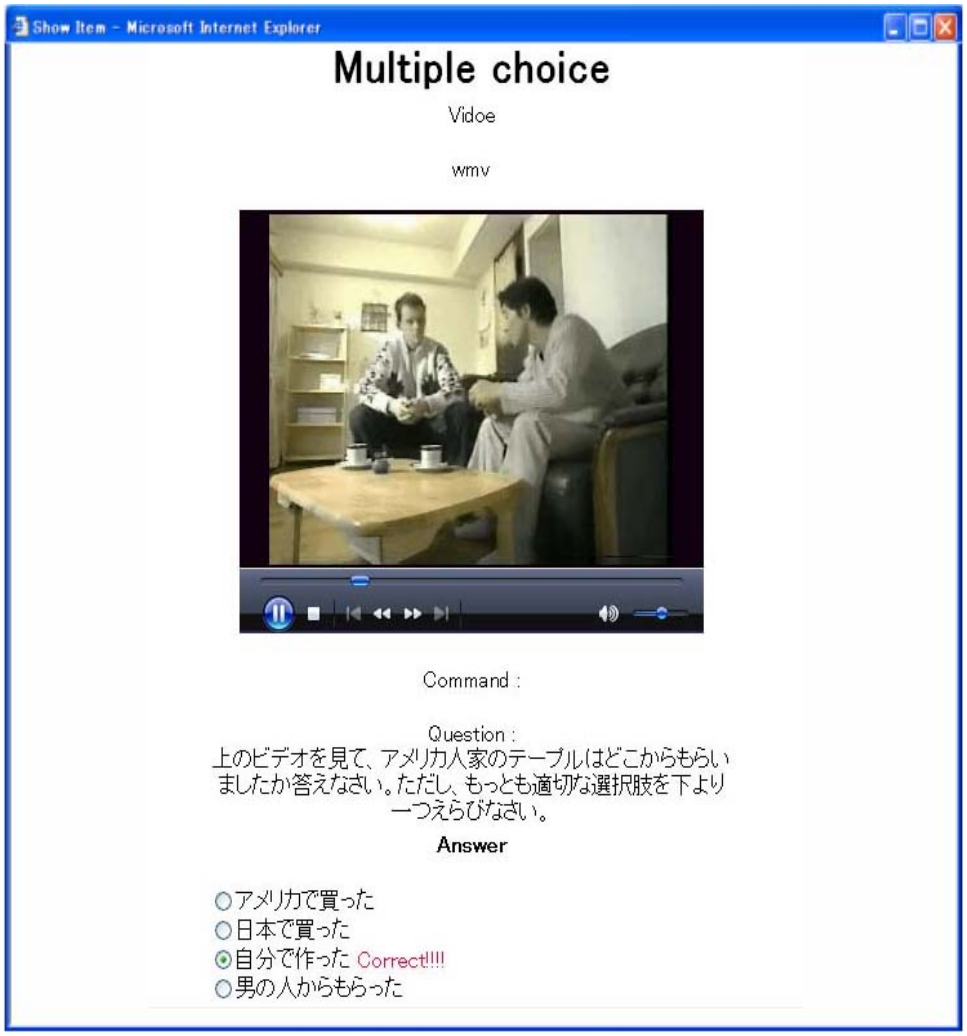
# Test construction support system



# Test Database

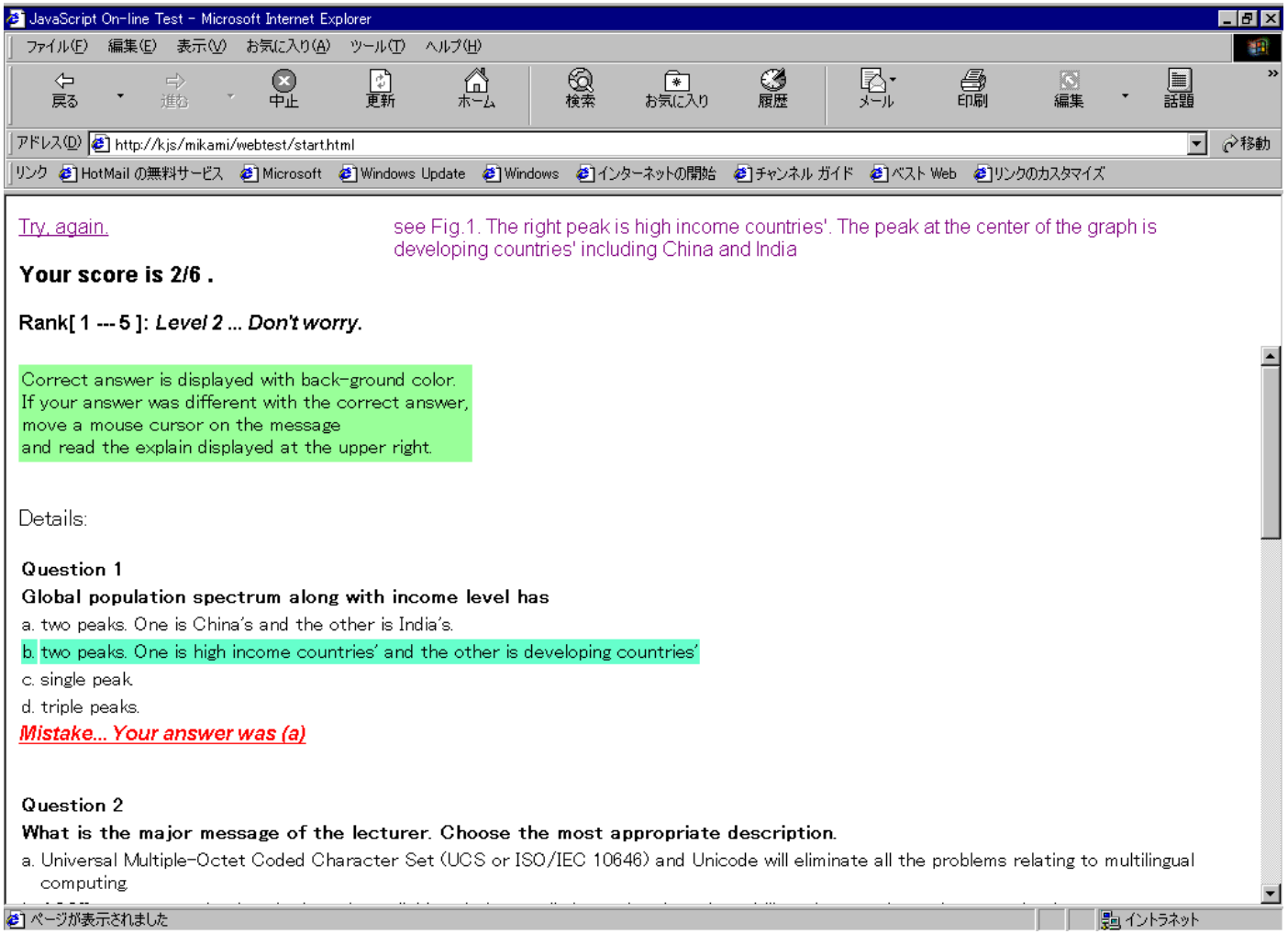
- Test information
  - Test ID
  - Test author
  - List of selected items
- Test history data
  - Examinee's answer process
    - Answers
    - Score on each item
    - Response-time

# Test delivery system

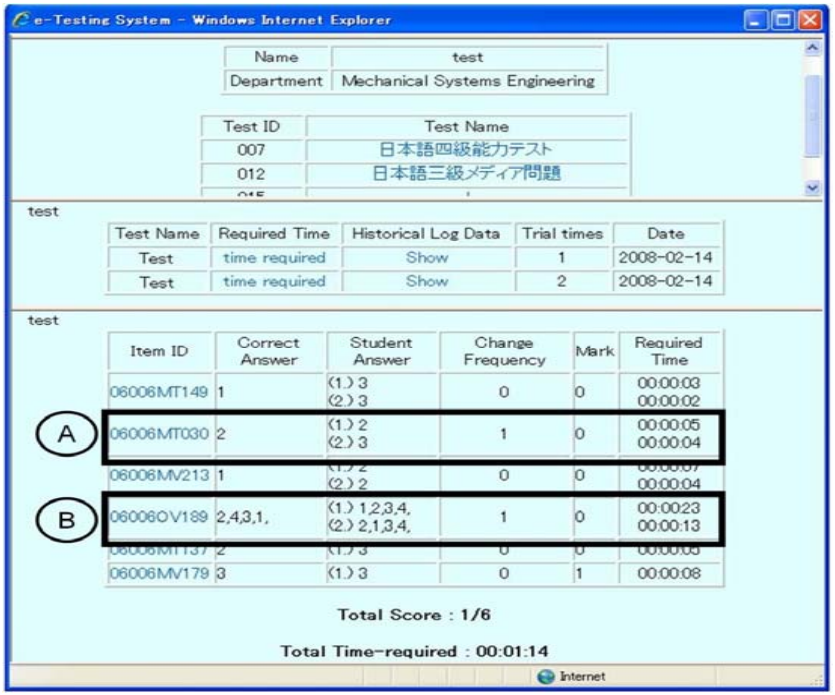



- Test distribution and answer collection can be automatically performed
- Item can be randomly selected
- Examinees in distance places can be evaluated

# Feed back for students

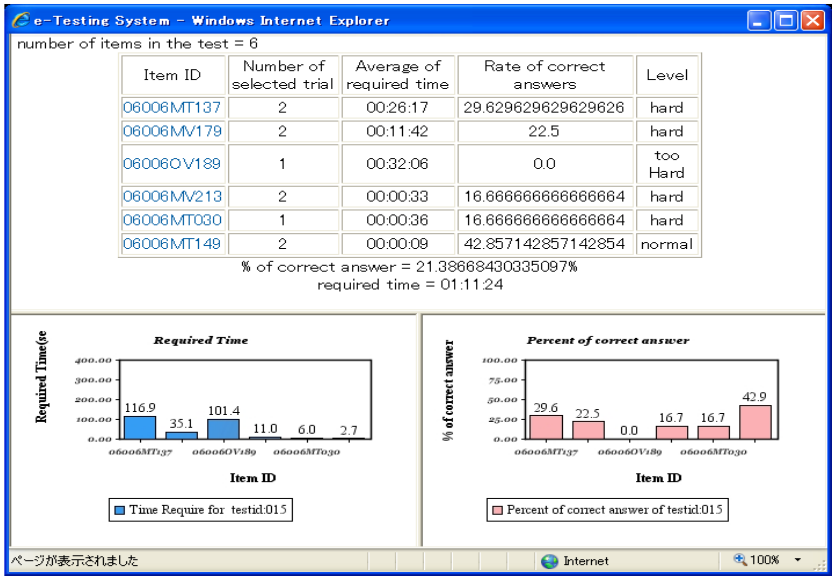


# Data analysis system



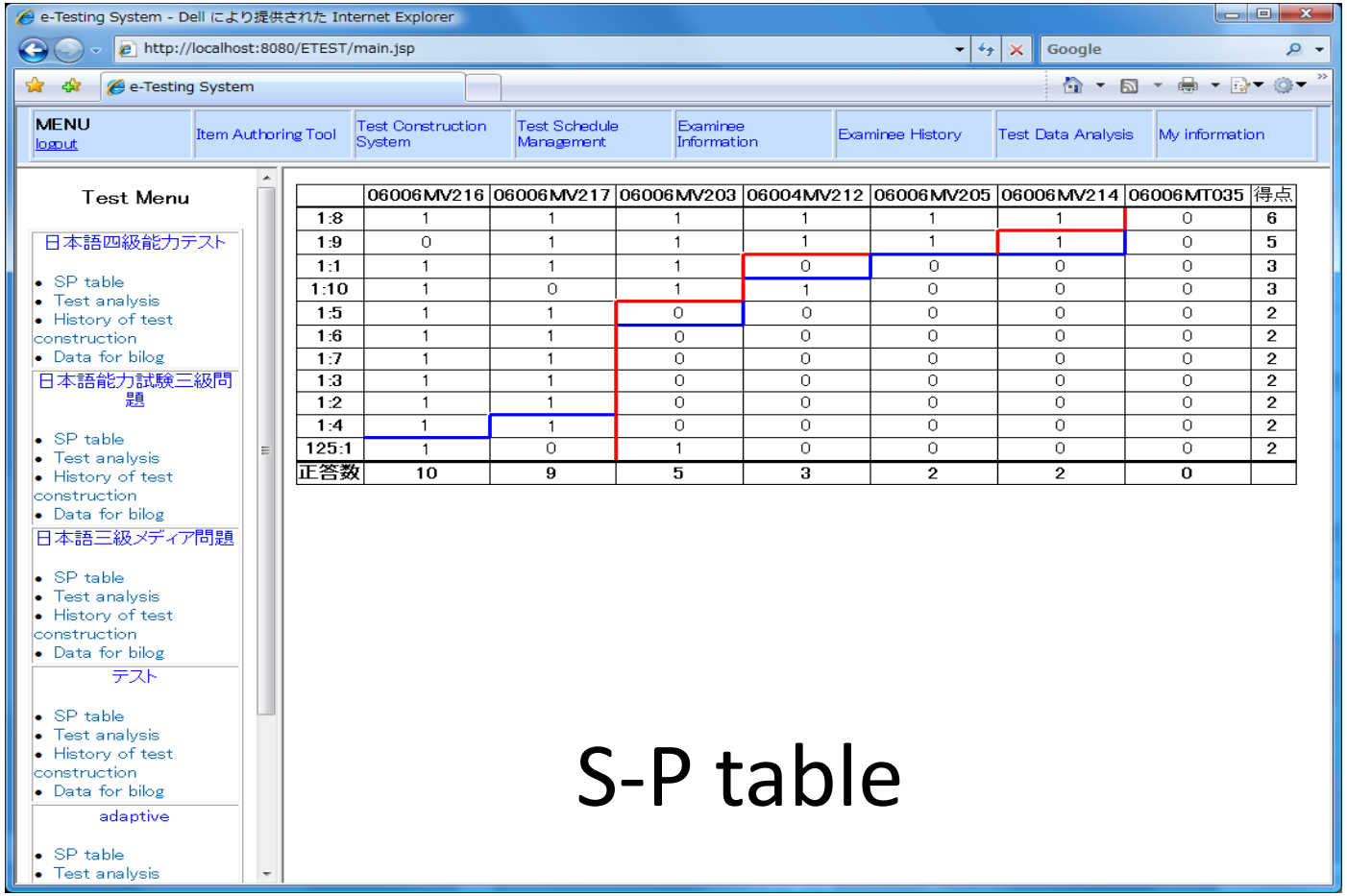

- Analysis of examinee
  - Item score
  - Test score
  - Item response-time
  - Answer changing history
  - Review history
  - Average test score
  - Average response-time

# Data analysis system




- Analysis of test.
  - Rate of correct answers
  - Average item response-time
  - Average answer changing frequency of item
  - Test score distribution
  - Test response-time distribution.

# Data analysis system

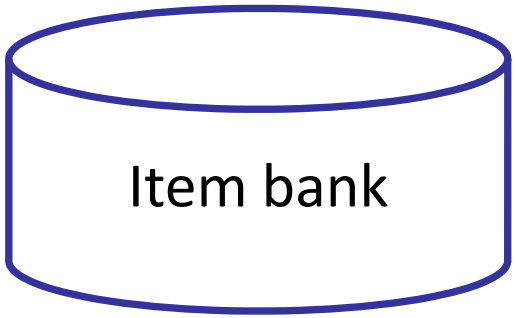



| | 06006MV216 | 06006MV217 | 06006MV203 | 06004MV212 | 06006MV205 | 06006MV214 | 06006MT035 | 得点 |
|---|---|---|---|---|---|---|---|---|
| 1:8 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 6 |
| 1:9 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 5 |
| 1:1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 3 |
| 1:10 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 3 |
| 1:5 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 1:6 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 1:7 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 1:3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 1:2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 1:4 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 125:1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| 正答数 | 10 | 9 | 5 | 3 | 2 | 2 | 0 | |

S-P table

# Two examples of advanced technologies

# 4. Automated test construction system




Select items

Automated test construction system

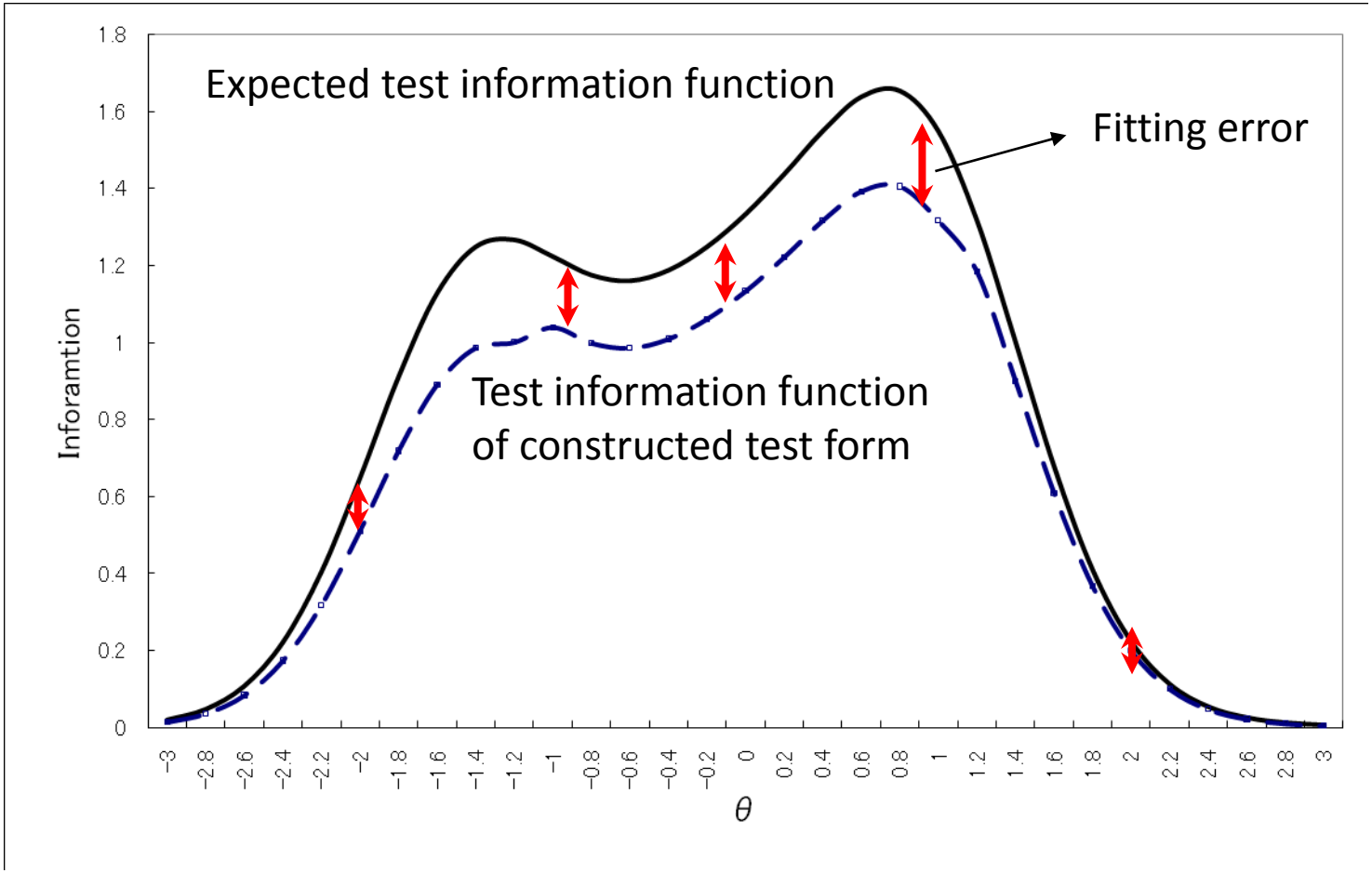
Test satisfying constraints


Constraints

Example :
- Average score 60 scores
- Response-time 30 minutes
- Number of items from subject A 10 items

# Test constraints setup interface

# Test information function constraint in multiple test forms construction

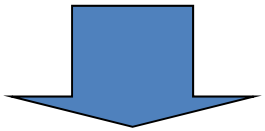


# Previous methods

Previous multiple test construction methods

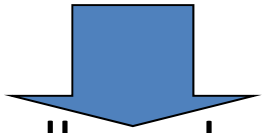
Simultaneous test construction（Adema,1990）

Big-Shadow Test (Linden, 2005)



Trade-off between the fitting errors differences and computational time

- Guarantee the maximum number of constructed test forms
  - Armstrong and Belov (2006)



- Overlapping item is not allowed
- Interrupt effective use of an item bank

# The fastest method

Pokpong Songmuang, Maomi Ueno, "Bees Algorithm for Construction of Multiple Test Forms in e-Testing," *IEEE Transactions on Learning Technologies*, 27 2010,

$$O(c \cdot m! \cdot 2^f) \longrightarrow O(c \cdot m! + 2^f)$$

Traditional Methods → Proposed Method

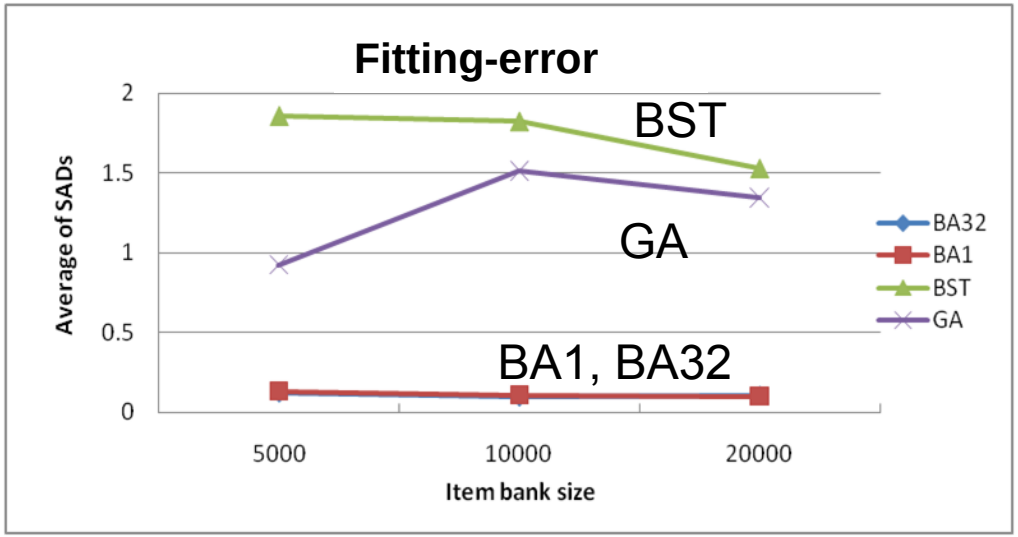
where

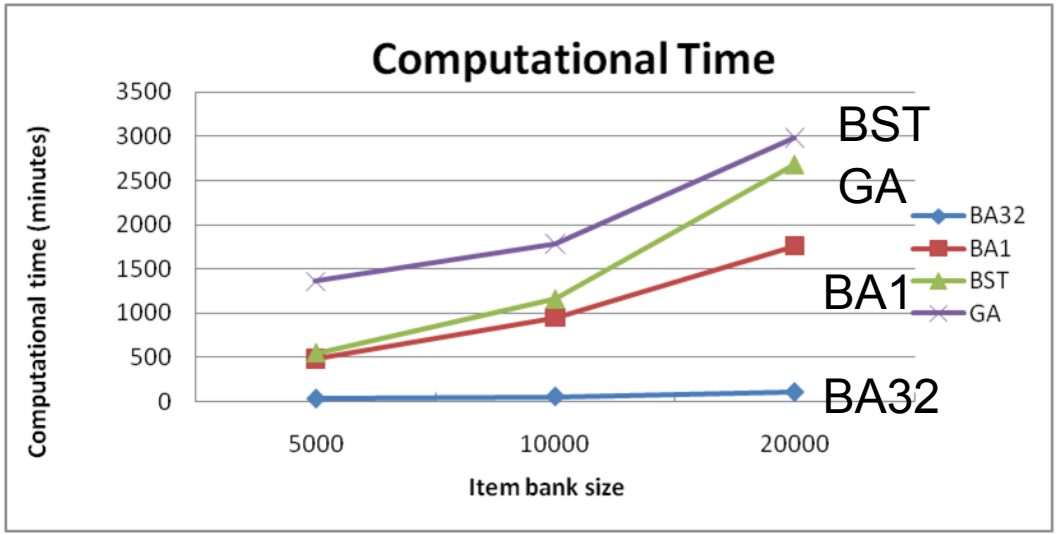
$c$ : the number of test constraints

$m$ : the number of items in an item bank

$f$ : the number of constructed test forms that satisfy all test constraints

# Experiments methods

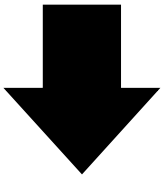



Fitting error : Summary of absolute differences of test information function

# 5. Adaptive Testing

The theoretical basis for most adaptive testing systems is the Item Response Theory (IRT).



Problems

1. Local independence assumption
2. Fisher information paradox
3. Only items with high discrimination parameter tend to be selected in a test.
4. Trade-off between prediction accuracy and computational costs

# Ueno ICALT 2010

# Adaptive Testing based on Decision Tree

$$p(u_i \mid u_1, u_2, \cdots, u_{i-1}) = \frac{1}{1+\exp(-L_i)}, \qquad (10)$$

where

$$L_i = \left[\log \frac{p(u_i = 1 \mid u_1, u_2, \cdots, u_{i-1})}{p(u_i = 0 \mid u_1, u_2, \cdots, u_{i-1})}\right]^{u_i} \left[\log \frac{p(u_i = 0 \mid u_1, u_2, \cdots, u_{i-1})}{p(u_i = 1 \mid u_1, u_2, \cdots, u_{i-1})}\right]^{(1-u_i)}$$

$$I(u_k \mid \theta) = \sum_{u_k=0,1} \int_\theta p(u_k \mid \theta) p(\theta) \log \frac{p(u_k \mid \theta)}{p(u_k)} d\theta. \qquad (12)$$

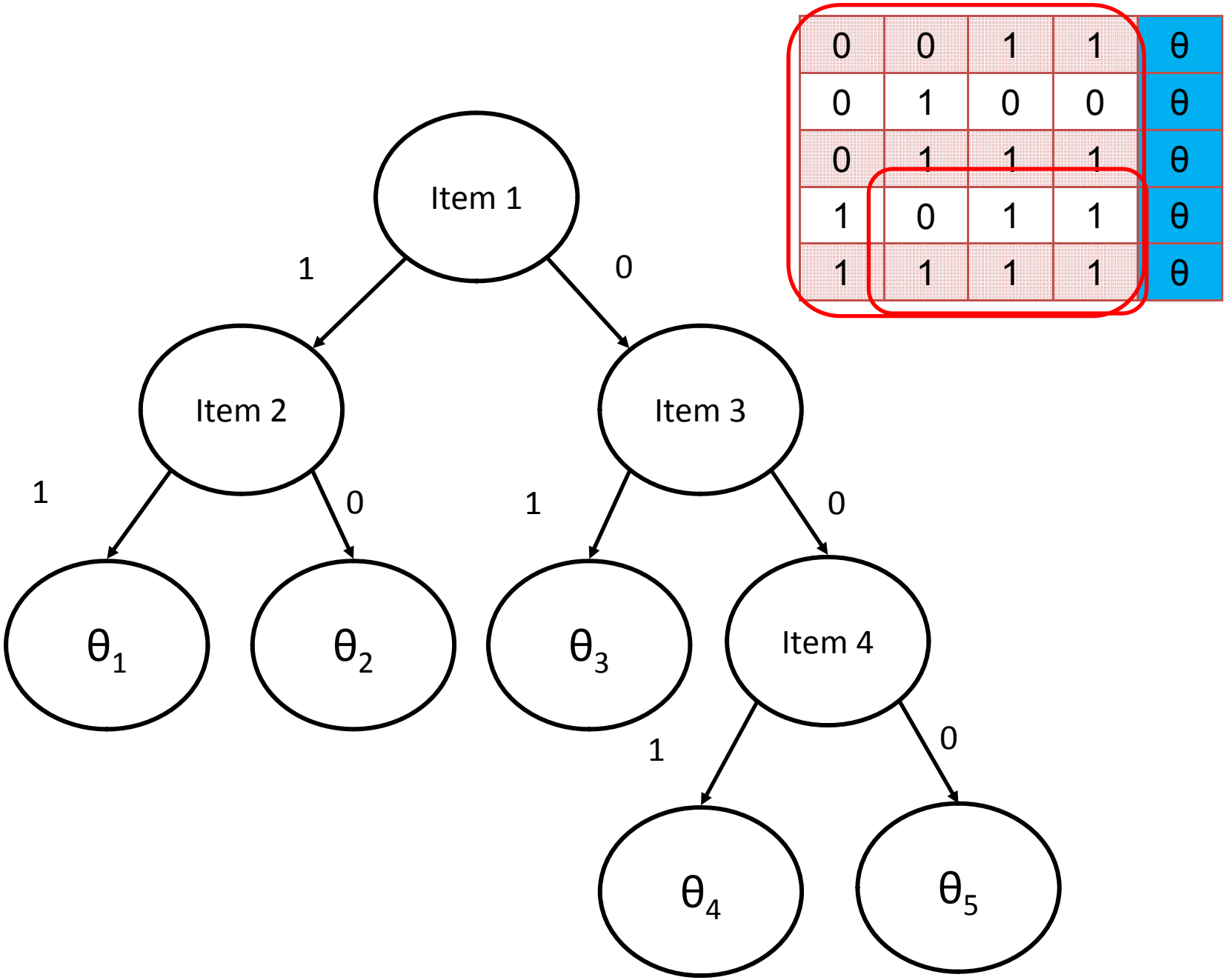

0 0 1 1 θ
0 1 0 0 θ
0 1 1 1 θ
1 0 1 1 θ
1 1 1 1 θ
Item 1
1
0
Item 2
Item 3
1
0
1
0
θ1
θ2
θ3
Item 4
1
0
θ4
θ5

TABLE I

COMPARISON OF ITEM SELECTION METHODS BASED ON LOCAL INDEPENDENT DATA

| | MSE (SD) | presented items (SD) | Time (SD) | High $a_i$ items frequency (SD) |
|---|---|---|---|---|
| Item Banks $I$=100 | | | | |
| FI | 1.18(0.56) | 32.41(25.52) | 1.54(1.39) | 168.25(48.72 ) |
| GI | 1.19(0.55) | 31.70(24.33) | 2.88(1.08) | 163.58(53.07 ) |
| LWI | 1.19(0.54) | 34.70(24.83) | 1.95(1.05) | 172.25(57.34 ) |
| EPWI | 1.16(0.55) | 25.96(23.46) | 4.55(3.42) | 131.52(74.78) |
| DT | 0.6934(0.27) | 18.57(11.51) | 1.08(3.24) | 96.38(73.13) |
| Item Banks $I$=500 | | | | |
| FI | 1.65(0.72) | 176.53(146.54) | 1.23(1.58) | 199.17(40.68) |
| GI | 1.69(0.70) | 176.29(134.18) | 9.37(1.61) | 198.41(45.69 ) |
| EPWI | 1.59(0.68) | 83.78(101.48) | 37.97(35.01) | 127.08(68.56 ) |
| DT | 0.50(0.20) | 66.84(107.36) | 1.38(1.88) | 41.08(48.45) |
| Item Banks $I$=1000 | | | | |
| FI | 1.50(0.70) | 400.16(265.50) | 1.61(2.18) | 218.87(41.22 ) |
| GI | 1.52(0.70) | 396.73(252.79) | 17.98(1.76) | 221.43(44.31 ) |
| LWI | 1.56(0.70) | 409.58(273.45) | 7.89(1.48) | 226.88(44.62) |
| EPWI | 1.60(0.65) | 168.68(212.97) | 133.11(143.65) | 121.26(67.77) |
| DT | 0.45(0.17) | 57.75(141.87) | 1.13(2.74) | 25.57(39.23 ) |

TABLE II

COMPARISON OF ITEM SELECTION METHODS BASED ON A NETWORK STRUCTURE DATA

| | MSE (SD) | presented items (SD) | Time(SD) | High $a_i$ items frequency (SD) |
|---|---|---|---|---|
| Item Banks $I$=100 | | | | |
| FI | 1.53 (0.81) | 43.05 (24.82) | 5.98 (23.47) | 241.08(64.66) |
| GI | 1.54 (0.82) | 43.26 (23.84) | 5.65 (12.86) | 242.07(69.36) |
| LWI | 1.56 (0.82) | 45.76 (25.05) | 4.86 (15.21) | 259.06(67.41) |
| EPWI | 1.4171 (0.78) | 29.20 (19.63) | 7.13 (10.98) | 157.88(95.19) |
| DT | 0.8787 (0.36) | 23.36 (192.85) | 1.32 (5.54) | 136.88(127.90) |
| Item Banks $I$=500 | | | | |
| FI | 1.73(0.92) | 191.49(124.91) | 6.29(22.84) | 240.40(35.26) |
| GI | 1.74(0.95) | 194.78(127.90) | 19.11(19.53) | 223.96(28.97) |
| LWI | 1.8885(0.9083) | 203.64(122.02) | 8.36(5.22) | 252.94(16.64) |
| EPWI | 1.60(0.86) | 72.50(63.56) | 46.37(79.16 ) | 80.78 (61.12) |
| DT | 0.78(0.32) | 19.24(159.03) | 1.31(3.40) | 28.89(65.75) |
| Item Banks $I$=1000 | | | | |
| FI | 2.11(0.93) | 484.85(286.57) | 11.20(44.35 ) | 321.10(45.57) |
| GI | 2.05(0.97) | 467.15(302.96) | 22.19(78.16 ) | 308.71(45.77) |
| LWI | 2.09(0.93) | 488.00(289.19) | 9.81(26.30 ) | 320.61(43.53) |
| EPWI | 1.79(0.87) | 95.07(164.09) | 83.88(120.52) | 103.57(113.62) |
| DT | 0.83(0.35) | 18.84(159.15) | 3.64(2.41) | 19.72(70.84) |

# 6. Practice of e-Testing in Japan

- Japan has more than 300 national examinations
- Information-technology Promotion Agency (IPA) provides *Information-technology engineers examination* which is the largest national examination in Japan (600,000 examinees/year)
- IPA plans to transfer paper-based examinations of all of national examinations to Internet-based examinations in 2011.
- The design was committed by our lab.

# 7. Conclusions

- E-Testing has unique key technologies.
- The demand for e-Testing has grown in society.
- Problem:  lack of experts in e-Testing .