

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

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

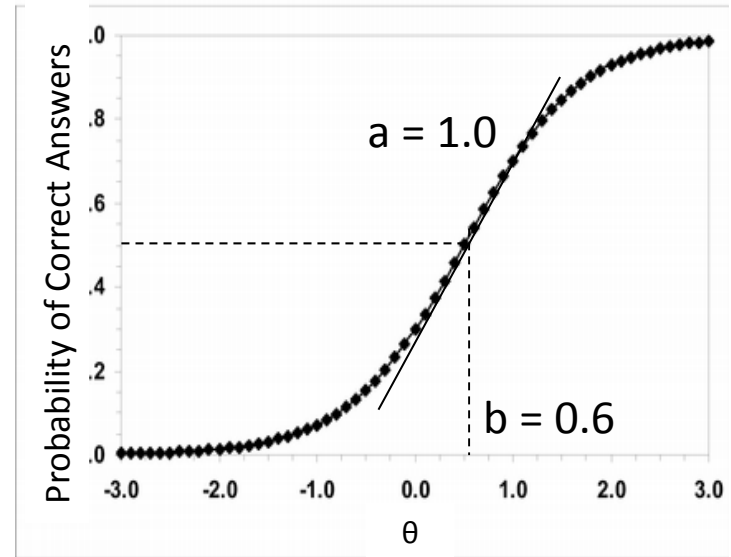
where

- θ_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

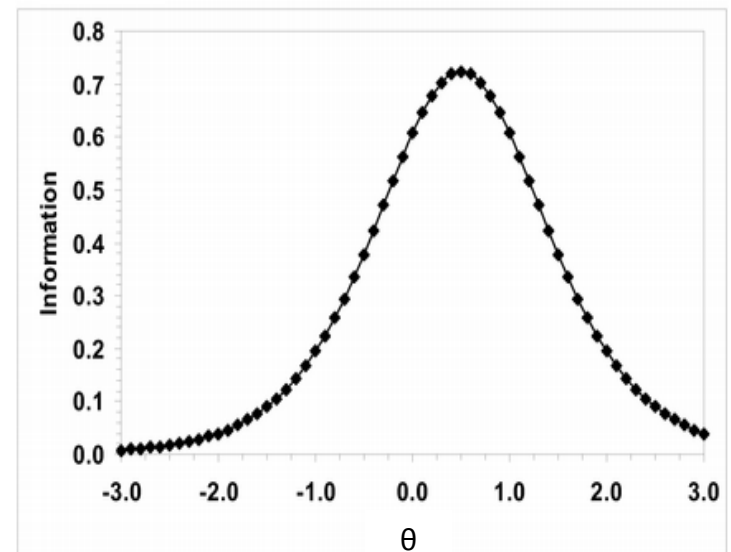
Item information function is

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

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

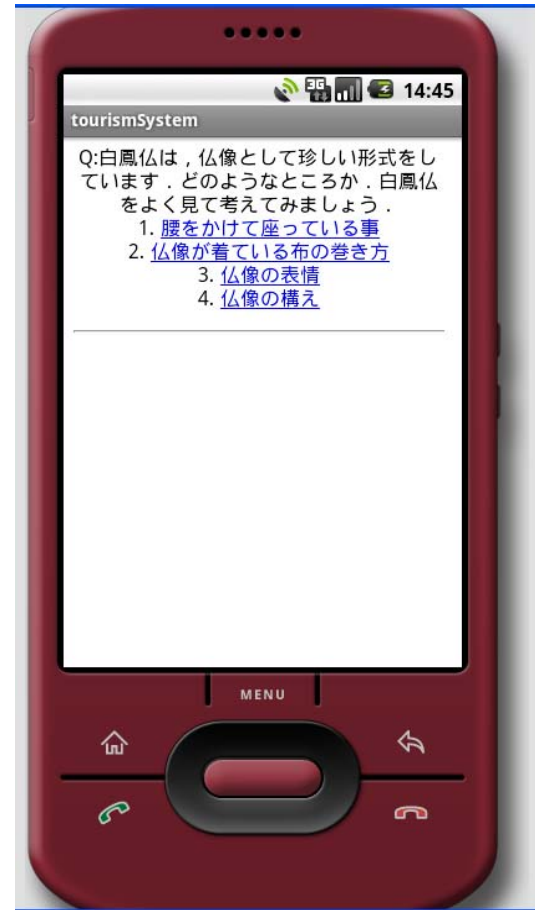


Input the name of a disease and the degree as parameters for Patient Robot

Mobile Adaptive Testing

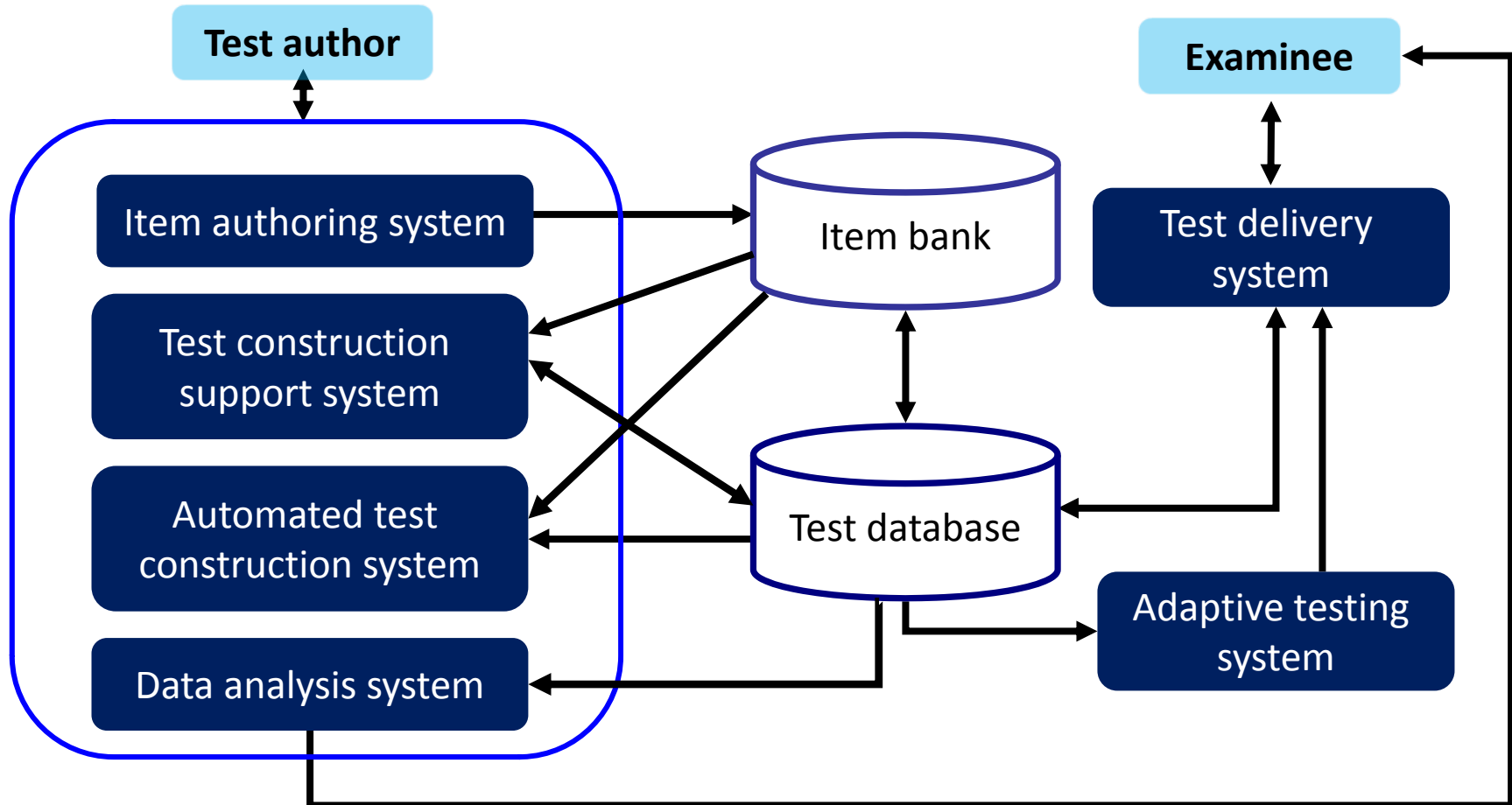


Navigation



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)

Internet Explorer window showing the 'Show Item' page for item ID 06006MPA093.

Item information

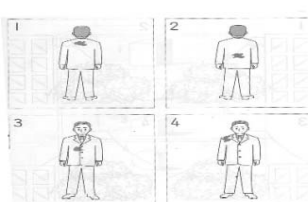
Subject Name	言語
Area Name	日本語テスト
Type	Multiple choice
Media	Picture & Audio
Item number	0093
Item ID	06006MPA093

Number if used in test	Average time required	Minimum time required	Maximum time required	Rate of correct answers	Level
1	00:01:01(0.0)	00:00:01	00:10:32	0.7959 (0.403028931799377)	no data

Multiple choice

Audio
mp3

Picture
jpg



Command :
問題をきいて、こたえなさい。

Question :
Answer

● 1 Correct!!!!
○ 2
○ 3
○ 4

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

Multiple choice

Video

Multiple choice

Video

Multiple choice

Video

Command :

Question :

上のビデオを見て、アメリカ人の家ではどこから来たか、正しいか答えなさい。たぶん、もっとも適切な選択肢を下より一つえらびなさい。

Answer

○アメリカで買った
○日本で買った
○自分で作った Correct!!!
○男の人からもらった

06006MT063	00:00:25
06006MT068	00:00:13
06006MT069	00:00:09
06006MT072	00:00:28
06006MT074	00:00:30
06006MT077	00:00:10
06006MT087	00:00:08

8080 - e-Testing Construction System with tools - Microsoft Internet Explorer

e-Testing Construction System with tools

Test Name	test
Test ID	078
Subjects	1
Definition	ttt
Expect Required Time	now showing in predictive response time 120 seconds submit

Author : test2
Invited Co-Authors Pokpong

10/17/2007 11:50:46

Reload

number of items in the test = 13
number of test-author(s) = 2

Item ID	Number of selected trial	Average of required time	Rate of correct answers
06006MPA102	21	00:00:33	0.8142857142857143
06006MT033	40	00:00:15	0.673913043478261
06006MT086	23	00:00:16	0.632183908045977
06006MT062			

Todays Date: 10/17/2007 11:50:48

agent:agent recommends 06006MT068

test2:item please

agent:test efficiency is improved after delete 06006MPA100

agent:test efficiency is improved after delete 06006MT037

agent:test efficiency is improved after add 06006MT089

agent:agent recommends 06006MT089

test2:item please

agent:test efficiency is decreased after delete 06006MT089

agent:test efficiency is improved after add 06006MT049

agent:agent recommends 06006MT049

test2:item please

agent:test efficiency

Stop Start

Submit

インターネット

Predictive Response Time Distribution

$\alpha=1.29, \beta=11.1, sd=0.289, mean=277.0$

Predictive Score Distribution

Percent Correct Answer, mean = 0.79

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

Show Item - Microsoft Internet Explorer

Multiple choice

Vidoe
wmv



Command :

Question :

上のビデオを見て、アメリカ人家のテーブルはどこからもらいましたか答えなさい。ただし、もっとも適切な選択肢を下より一つえらびなさい。

Answer

- ☐ アメリカで買った
- ☐ 日本で買った
- ☒ 自分で作った **Correct!!!!**
- ☐ 男の人からもらった

- 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

JavaScript On-line Test - Microsoft Internet Explorer

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 進む 中止 更新 ホーム 検索 お気に入り 履歴 メール 印刷 編集 話題

アドレス(D) <http://kjs/mikami/webtest/start.html> 移動

リンク [HotMail の無料サービス](#) [Microsoft](#) [Windows Update](#) [Windows](#) [インターネットの開始](#) [チャンネル ガイド](#) [ベスト Web](#) [リンクのカスタマイズ](#)

[Try again.](#) see Fig. 1. The right peak is high income countries'. The peak at the center of the graph is developing countries' including China and India

Your score is 2/6 .

Rank[1 --- 5]: Level 2 ... Don't worry.

Correct answer is displayed with back-ground color.
If your answer was different with the correct answer,
move a mouse cursor on the message
and read the explain displayed at the upper right.

Details:

Question 1
Global population spectrum along with income level has

- a. two peaks. One is China's and the other is India's.
- b. two peaks. One is high income countries' and the other is developing countries'**
- c. single peak
- d. triple peaks.

Mistake... Your answer was (a)

Question 2
What is the major message of the lecturer. Choose the most appropriate description.

- a. Universal Multiple-Octet Coded Character Set (UCS or ISO/IEC 10646) and Unicode will eliminate all the problems relating to multilingual computing

ページが表示されました

インターネット

Data analysis system

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

e-Testing System - Windows Internet Explorer

Name	test
Department	Mechanical Systems Engineering

Test ID	Test Name
007	日本語四級能力テスト
012	日本語三級メディア問題

test

Test Name	Required Time	Historical Log Data	Trial times	Date
Test	time required	Show	1	2008-02-14
Test	time required	Show	2	2008-02-14

test

Item ID	Correct Answer	Student Answer	Change Frequency	Mark	Required Time
06006MT149	1	(1.) 3 (2.) 3	0	0	00:00:03 00:00:02
(A) 06006MT030	2	(1.) 2 (2.) 3	1	0	00:00:05 00:00:04
06006MV213	1	(1.) 2 (2.) 2	0	0	00:00:07 00:00:04
(B) 06006OV189	2,4,3,1,	(1.) 1,2,3,4, (2.) 2,1,3,4,	1	0	00:00:23 00:00:13
06006MT137	2	(1.) 3	0	0	00:00:03
06006MV179	3	(1.) 3	0	1	00:00:08

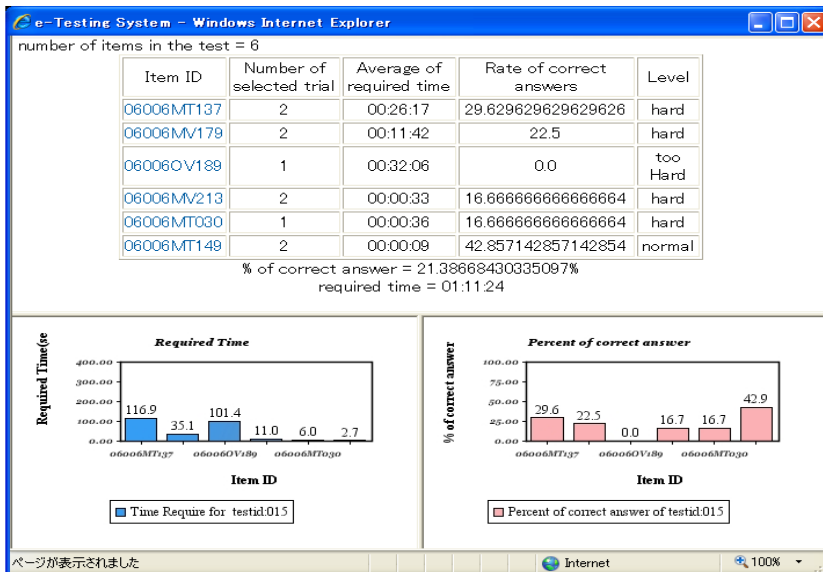
Total Score : 1/6

Total Time-required : 00:01:14

Internet

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

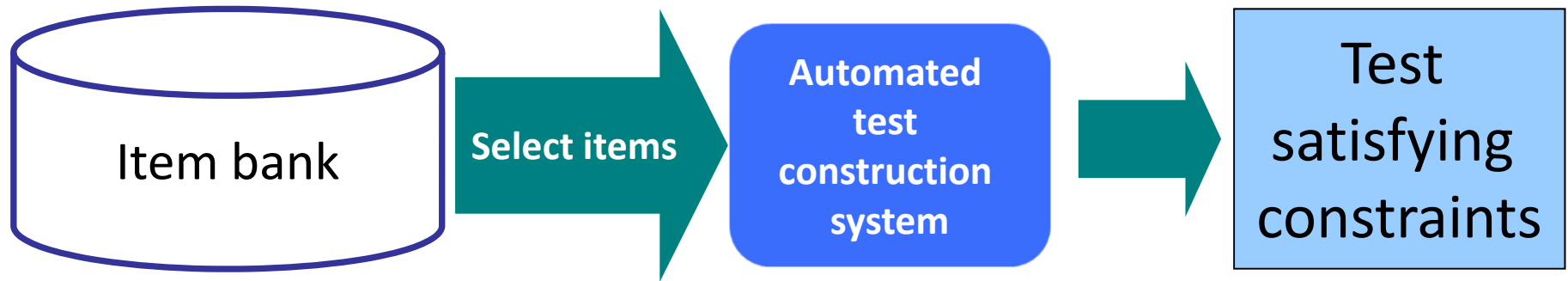
The screenshot shows a web browser window titled "e-Testing System - Dell により提供された Internet Explorer". The address bar shows "http://localhost:8080/ETEST/main.jsp". The browser window contains a navigation menu with the following items: MENU, Item Authoring Tool, Test Construction System, Test Schedule Management, Examinee Information, Examinee History, Test Data Analysis, and My information. The "Test Menu" is expanded, showing a list of test categories and their associated actions (SP table, Test analysis, History of test construction, Data for bilog). The categories are: 日本語四級能力テスト, 日本語能力試験三級問題, 日本語三級メディア問題, and adaptive. The "Test Data Analysis" section displays an S-P table with the following data:

	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



The screenshot shows the 'Automated Test Construction System' interface. It includes fields for 'Date' (2008/08/09), 'User', 'Construction Method' (Bee Colony Algorithm), and 'Keyword'. There are also input fields for 'Number of Test Items' (p1, p2, p3, p4, p5) and 'Parameter a (Lower bound)'. A table on the right lists 'Subjects', 'Area', 'Question Types', and 'Number of Items'. The 'Subjects' table shows 'Japanese Proficiency Test Level 1' with 'Reading' (1, 2, 3, 4, 5, 6, 7) and 'Writing' (1, 2, 3, 4, 5, 6, 7) sections. The 'Question Types' table shows 'True/False', 'Multiple Choice', 'Single Answer', and 'Sequencing'.

Subjects	Area	Question Types	Number of Items
Japanese Proficiency Test Level 1	Reading	1	True/False
		2	Multiple Choice
	Writing	3	True/False
		4	Multiple Choice
		5	Single Answer
		6	Multiple Choice
		7	Sequencing
		Listening	1
	2		Multiple Choice
	3		True/False
	4		Multiple Choice
	5		Single Answer
	6		Multiple Choice

Constraints

Example :

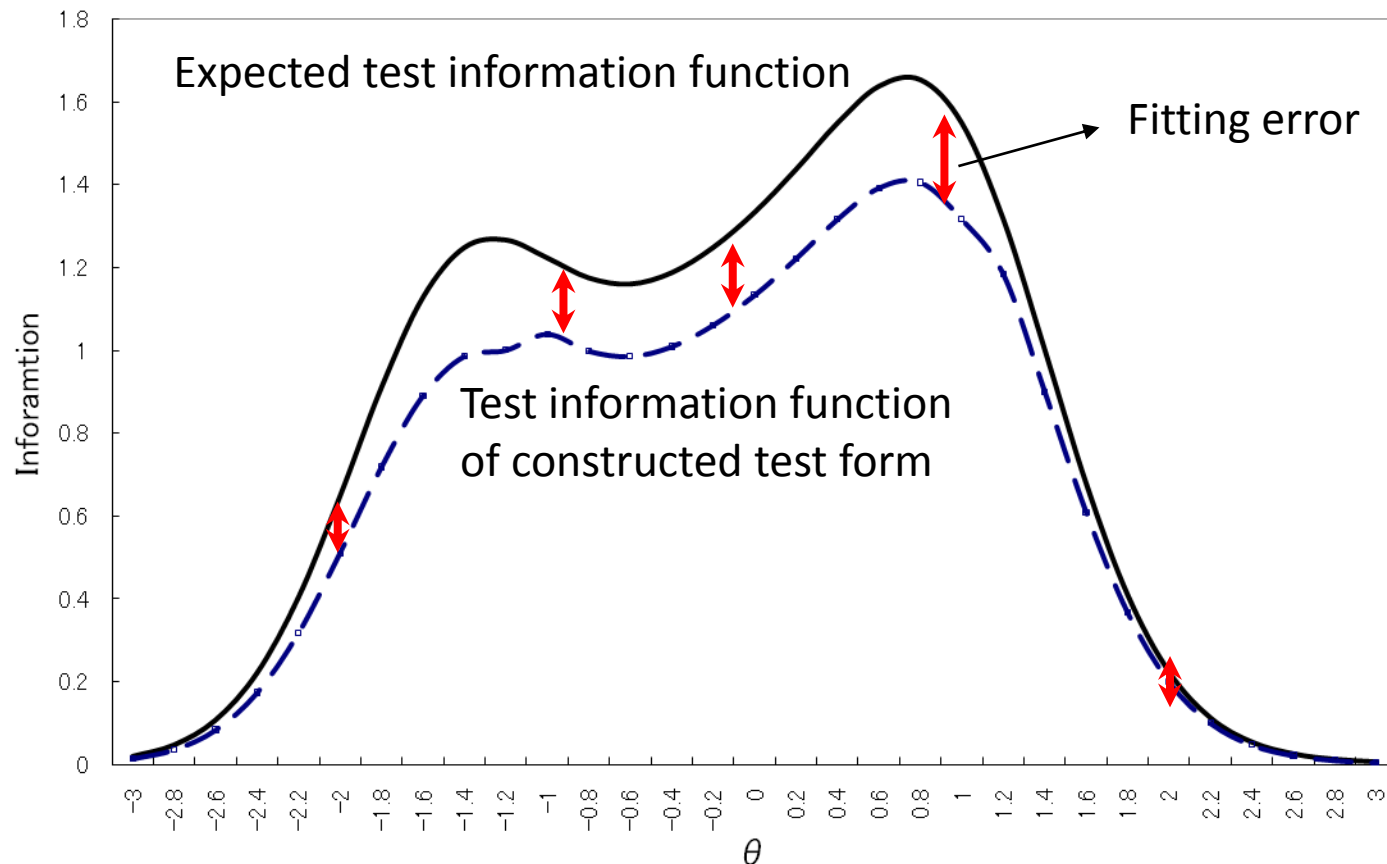
- Average score 60 scores
- Response-time 30 minutes
- Number of items from subject A 10 items

Test constraints setup interface

テスト構成支援システム									
ファイル(F) 編集(E) 表示(V) 履歴(I) ブックマーク(B) ウィンドウ(W) ヘルプ(H)									
作成日	2008/06/20								
Ver.									
構成方法	マニュアル ▼								
キーワード									
備考									
問題セット数									
项目实施回数 (下限値)									
テスト間類似度 (下限値)	(0~1)								
実質領域数 (下限値)	分野	低		~	高				
	大分類	低		~	高				
	中分類	低		~	高				
パラメータ (下限値)									
テスト情報量 (下限値)	(既定値=xxx)								
可否0									
可否テスト情報量 (下限値)	(既定値=xxx)								
正答率	平均	低		~	高				
	標準偏差	低		~	高				
所要時間 (秒)	平均	低		~	高				
	標準偏差	低		~	高				
合計問題数									
実行									

タイプ	分野	分野別抽出数 (下限値)	大分類	中分類	抽出数 (下限値)		
小問	ストラテジ系	☐	1 企業と法務	1 企業活動	☐		
				2 法務	☐		
			2 経営戦略	3 経営戦略マネジメント	☐		
				4 技術戦略マネジメント	☐		
				5 ビジネスインダストリ	☐		
			3 システム戦略	☐	6 システム戦略	☐	
					7 システム企画	☐	
					8 システム開発技術	☐	
					9 ソフトウェア開発管理技術	☐	
	マネジメント系	☐	4 開発技術	10 プロジェクトマネジメント	☐		
				11 サービスマネジメント	☐		
			5 プロジェクトマネジメント	12 システム監査	☐		
				13 基礎理論	☐		
			テクノロジー系	☐	6 サービスマネジメント	14 アルゴリズムとプログラミング	☐
						15 コンピュータ構成要素	☐
	7 基礎理論	16 システム構成要素			☐		
		17 システム構成要素			☐		
		18 ハードウェア			☐		
	8 コンピュータシステム	☐			19 ヒューマンインタフェース	☐	
					20 マルチメディア	☐	
					21 データベース	☐	
					22 ネットワーク	☐	
					23 セキュリティ	☐	
9 技術要素	☐						
中門	ストラテジ系	☐					
	マネジメント系	☐					
	テクノロジー系	☐					

Test information function constraint in multiple test forms construction

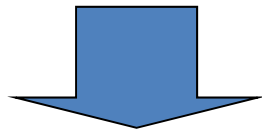


Previous methods

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

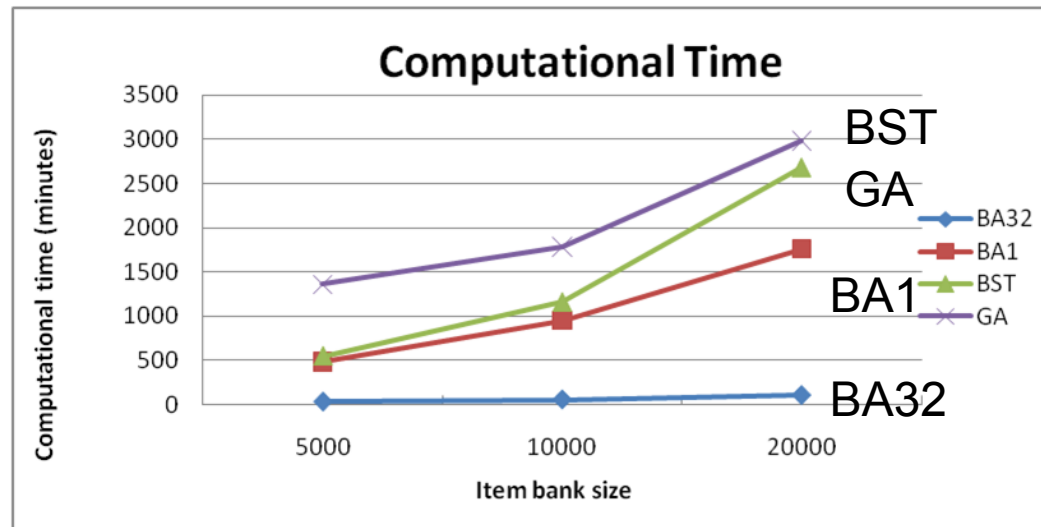
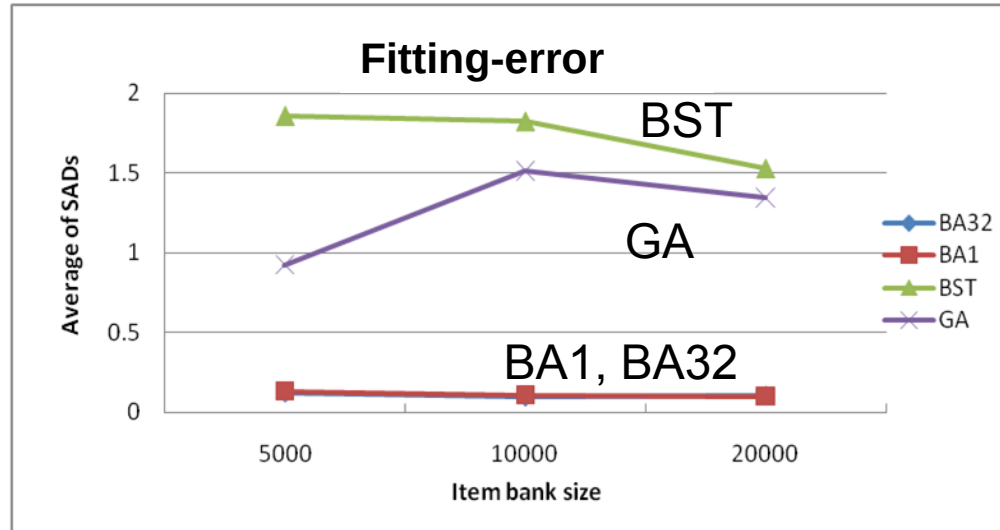
where

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

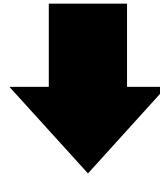
Experiments methods



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

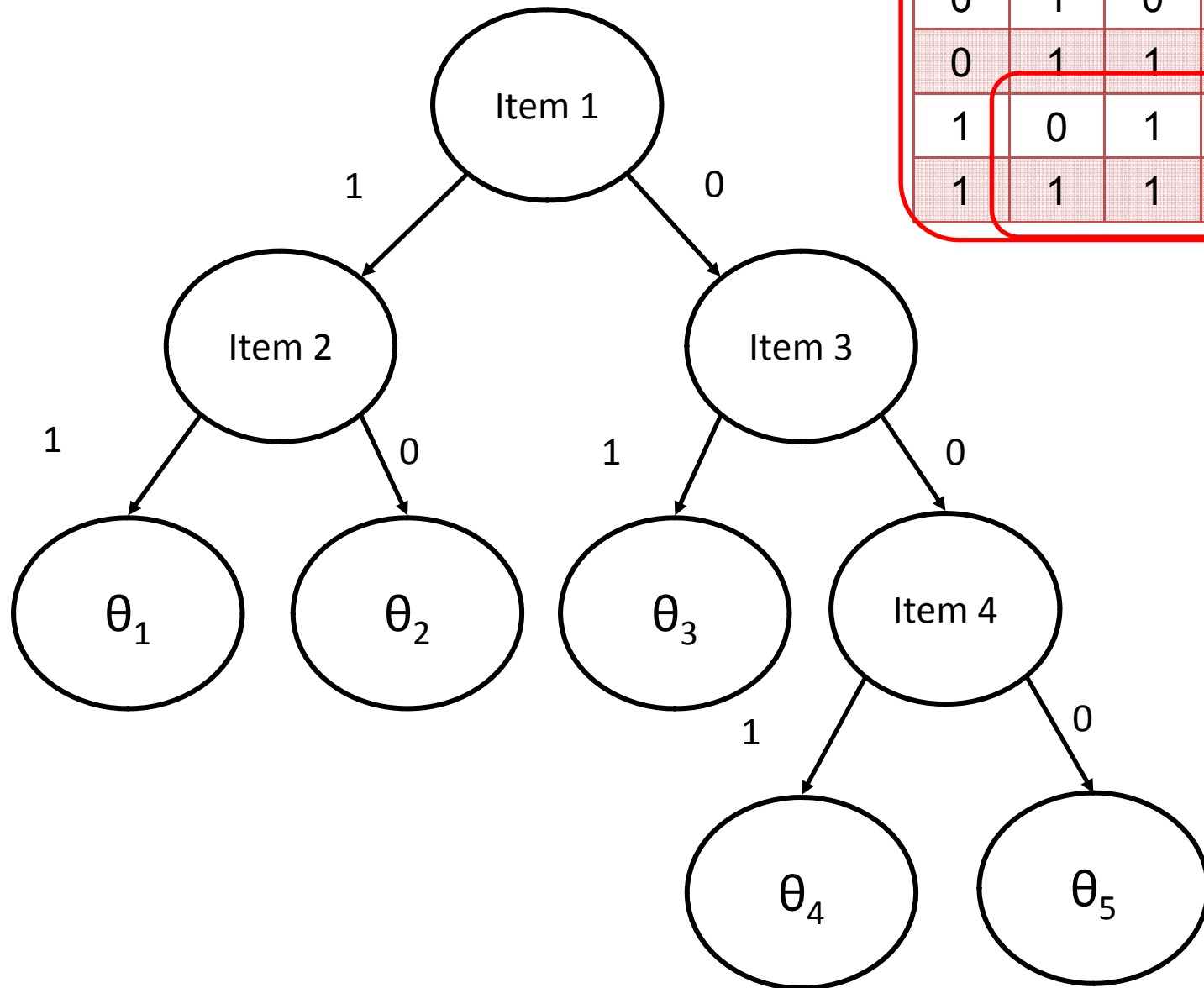
$$p(u_i \mid u_1, u_2, \dots, u_{i-1}) = \frac{1}{1 + \exp(-L_i)}, \quad (10)$$

where

$$L_i = \begin{bmatrix} \log \frac{p(u_i = 1 \mid u_1, u_2, \dots, u_{i-1})}{p(u_i = 0 \mid u_1, u_2, \dots, u_{i-1})} \end{bmatrix}^{u_i} \begin{bmatrix} \log \frac{p(u_i = 0 \mid u_1, u_2, \dots, u_{i-1})}{p(u_i = 1 \mid u_1, u_2, \dots, u_{i-1})} \end{bmatrix}^{(1-u_i)}$$



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



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

TABLE I
COMPARISON OF ITEM SELECTION METHODS BASED ON LOCAL INDEPENDENT DATA

	MSE (SD)	presented items (SD)	Time (SD)	High a_4 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_4 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 .