

The Development and Efficiency of the Computerized Adaptive Testing Version of The Test of Chinese as a Foreign Language (TOCFL-CAT)

Po-Hsi Chen¹ & Chia-Ling Hsu²

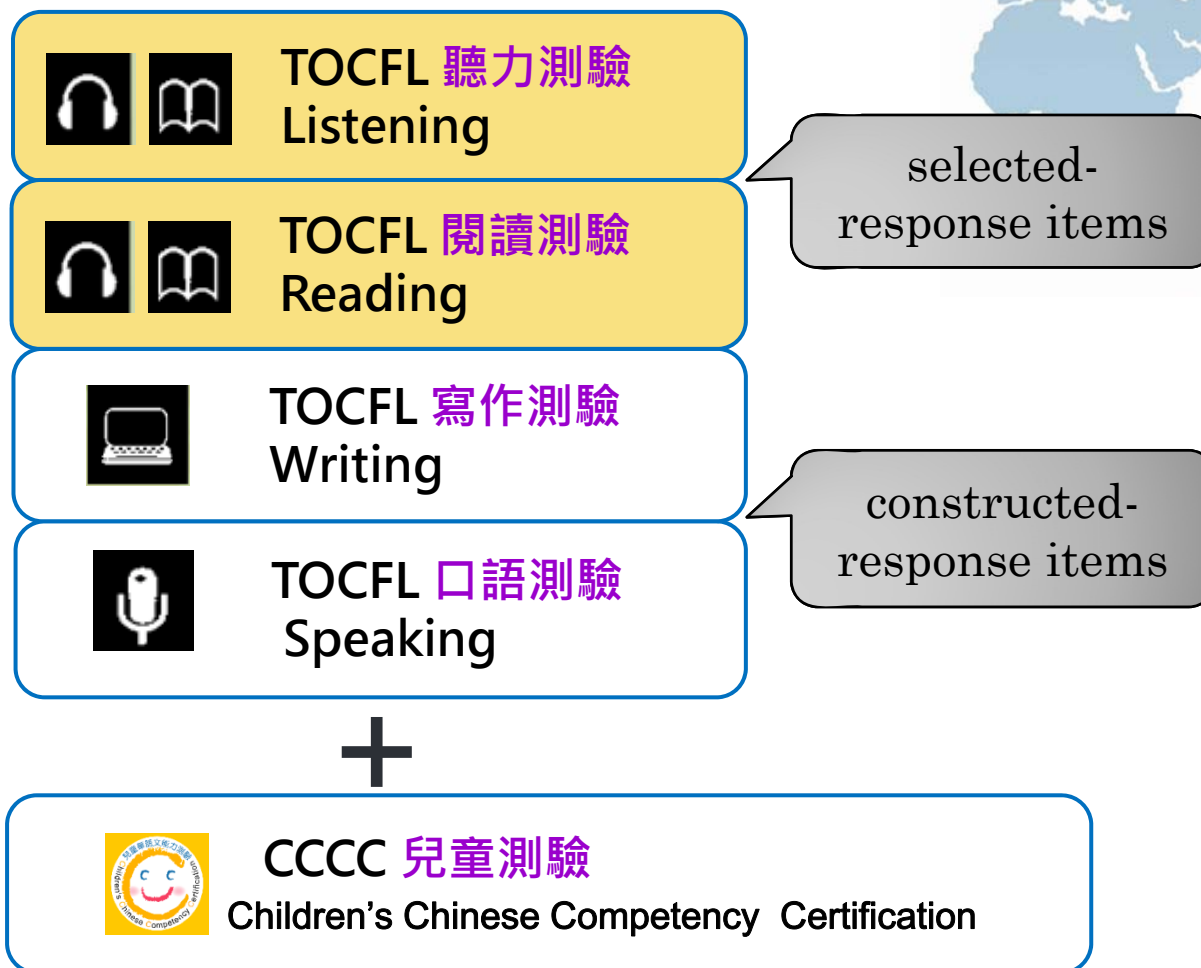
¹National Taiwan Normal University

²Hong Kong Institute of Educational



國家華語測驗推動工作委員會

Test of Chinese as a Foreign Language



TOCFL Worldwide - 25 Countries



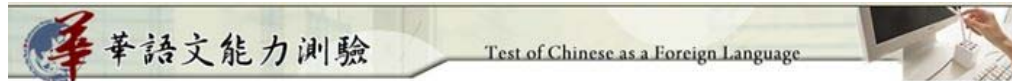
TOCFL Levels

TOCFL		CEFR
Band A	Level 1	A1
	Level 2	A2
Band B	Level 3	B1
	Level 4	B2
Band C	Level 5	C1
	Level 6	C2



TOCFL 聽力測驗

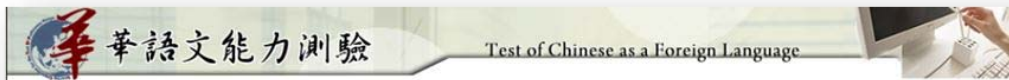
Listening



9)--(A)----D00167000. mp3



- ☐ (A)
- ☐ (B)
- ☐ (C)



11)--(A)----D00167100. mp3



- ☐ (A)
- ☐ (B)
- ☐ (C)

TOCFL 閱讀測驗

Reading

華語文能力測驗 Test of Chinese as a Foreign Language

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41-45 46-50

■ 已作答 ■ 已註記 □ 未作答

36) --(B)----



九月五日是小女孩的_____。

☐ (A) 生活
☒ (B) 生日
☐ (C) 星期日

Steering Committee for the Test Of Proficiency - Huayu



華語文能力測驗 Test of Chinese as a Foreign Language

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41-45 46-50

■ 已作答 ■ 已註記 □ 未作答

46-50) --(BEACF)----

昨天晚上我覺得很不舒服，(I)，所以很早就去睡覺了。今天早上起來，(II)。去看醫生，醫生說我感冒了，給了我一些藥。(III)要多休息，多喝水，才會快點好。這幾天的天氣一會兒熱，一會兒冷，(IV)。我要(V)，不要再感冒了。

	I	II	III	IV	V
(A) 還告訴我	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(B) 頭有點兒痛	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(C) 很容易生病	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(D) 覺得很舒服	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(E) 更不舒服了	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(F) 多注意自己的身體	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Steering Committee for the Test Of Proficiency - Huayu

TOCFL 口語測驗 Speaking



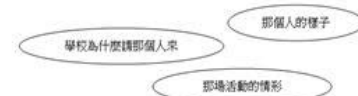
TOCFL 寫作測驗

Writing



寫作說明：

你參加了你們學校舉辦的一場活動，這場活動請來了一位有名的人。請你用李大明這個名字給你的朋友王天華寫一封信，告訴他這件事，這封信的內容必須有三個部分（※請注意：這三個部分的先後順序可以自己安排）：



注意事項：

- 寫作時間 40 分鐘。寫完就可以交卷。時間到了，電腦就自動跳到下一題，不能再修改內容。
- 請寫 250~350 個字。字數太多或是太少都會扣分。
- 寄信人和收信人必須用題目設定的姓名，**不可以**寫出自己的名字。
- 必須使用中式標點符號（，。：？！「」）。
- 請注意分段問題：
 - 如果全部寫成對話形式或是一行寫一句話，就 0 分。
 - 如果分段不適當也會扣分。

高階聽理解文

題目：Wang Daming's letter
 收信人：Wang Tianhua's letter
 內容：(The content of the letter is blurred in the screenshot)

寫作說明：

以上是王大明寄給好朋友林文惠的一封信，請你把自己當成林文惠，回信給王大明。信的內容必須包含以下三個部分：

- 自己或他人的相關經驗。
- 你對這件事的看法。
- 安慰或鼓勵對方。

注意事項：

- 信中不可以寫出自己的真實姓名。
- 可以自己決定上面三個寫作內容的順序。
- 文長 500-600 個字，過多過少都扣分。
- 請注意分段，並符合中式書信格式。
- 要用中式標點符號（，。：？！「」）。
- 每一題的寫作時間是 60 分鐘。



○ Why do we need TOCFL-CAT?

- Many test takers felt it difficult to choose the suitable test bands for them to take. In CAT, they don't need to choose.
- Too many items were needed to judge the examinees' proficiency level. CAT needs fewer items than non-adaptive testing in achieving the same degree of measurement precision.
- The stress of taking the test is different for examinees with different proficiency levels. Lower proficiency examinees would feel higher stress than higher proficiency examinees when taking the items with same difficult level.

測驗對象 Target Test Takers

Non-native Chinese speakers

等級 Level		CEFR	學時建議 Suggested Learning Hours	詞彙量建議 Suggested Vocabulary
Band A	入門級 Level 1	A1	120~	500
	基礎級 Level 2	A2	180~	1,000
Band B	進階級 Level 3	B1	360~	2,500
	高階級 Level 4	B2	480~	5,000
Band C	流利級 Level 5	C1	960~	8,000
	精通級 Level 6	C2		

聽力測驗 TOCFL LISTENING



	Band A	Band B , Band C
題數 number of items	50 items	
考試時間 test time	60 min (approx.)	
題型 item types	看圖回答 Picture Description 問答理解 Q & A Comprehension 對話理解 Dialogue Comprehension	對話聽力 Dialogue 段落聽力 Monologue



閱讀測驗 TOCFL READING



	Band A	Band B , Band C
題數 number of items	50 items	
考試時間 test time	60 min	
題型 item types	單句理解 Sentence Comprehension 看圖釋義 Picture Description 選詞填空 Gap Filling 完成段落 Paragraph Completion	選詞填空 Gap Filling 閱讀理解 Reading Comprehension



○ What is CAT?

- The test takers don't need to choose which bands of test to take.
- Testing system will choose the items for examinees automatically based on their answer in the prior items they took.
- If they answer an item correctly, the next item would become a difficult item. If they answer an item incorrectly, the next item would become an easy item.
- Most of the items are correspondent with the examinees' proficiency levels. The stress would be almost the same for all proficiency levels of examinees.



PURPOSE

- The aims of the study are to develop variable-length CAT algorithms for TOCFL listening and reading tests to provide the same measurement precision for examinees with different levels.
 - In simulation 1, we evaluated the efficiency of CAT by selecting items adaptively and non-adaptively.
 - In simulation 2, we evaluated performance of variable-length CAT by manipulating different constraints into item selection.



GENERATED DATA

- Examinees: 10,000 ~ $N(0,1)$
 - Covering three bands of test(A, B, and C) according to TOCFL's cut-off scores, and three bands consisted of 2,469, 5,452, and 2082 examinees, respectively.
- Item pool: 300 items drawn from TOCFL item bank
 - Including three bands of test(A, B, and C) in accordance with TOCFL's assembling criteria, each band composed of 50 items.



SIMULATION 1



- Independent variable: adaptive or non-adaptive
 - Non-adaptive (CBT version): examinees in each proficiency band are administered to the Band A, B, or C items.
 - Adaptive (fixed-length CAT version): 10,000 examinees administered to 30~50 items which are adaptively selected from 300 items.
- Dependent variable
 - standard errors of six levels of passing score in TOCFL



SIMULATION 2

- Item selection: maximum information (MI) method
- Termination: fixed-precision with standard error of 0.3
- Independent variables
 - Constraints: test type, test topic, and number of testlet items.
 - Maximum test length: 30, 35, 40, 45, and 50 item.
 - 40 conditions: 8 selection rules combined with 5 test length.

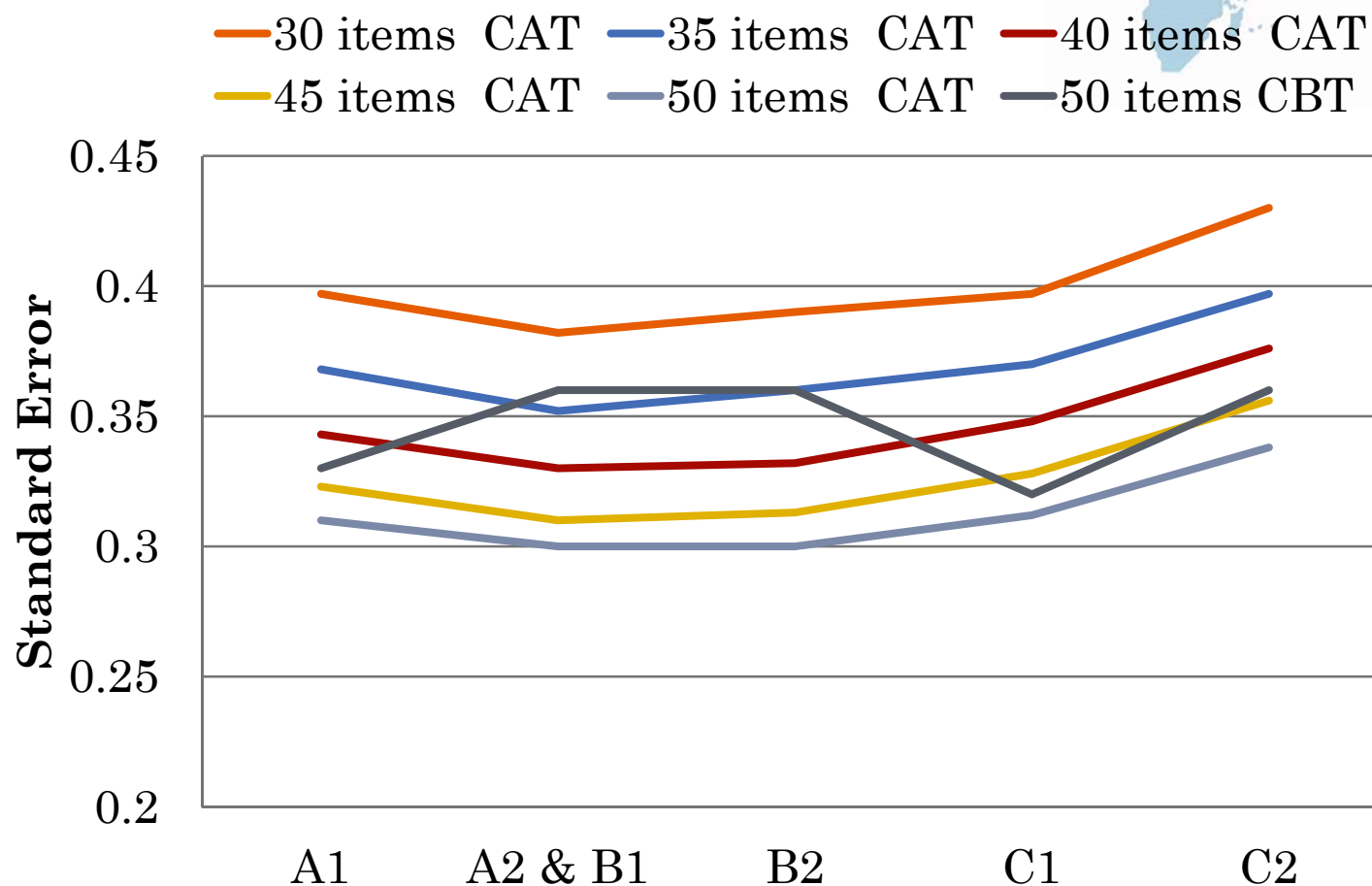
selection rules	maximum test length
• MI	• 30
• MI + type	• 35
• MI + topic	• 40
• MI + items	• 45
• MI + type + topic	• 50
• MI + type + items	
• MI + topic + items	
• MI + type + topic + items	

- Dependent variables:
 - Standard errors of the six levels of passing score in TOCFL.
 - Percentage of correctly classify (PCC).



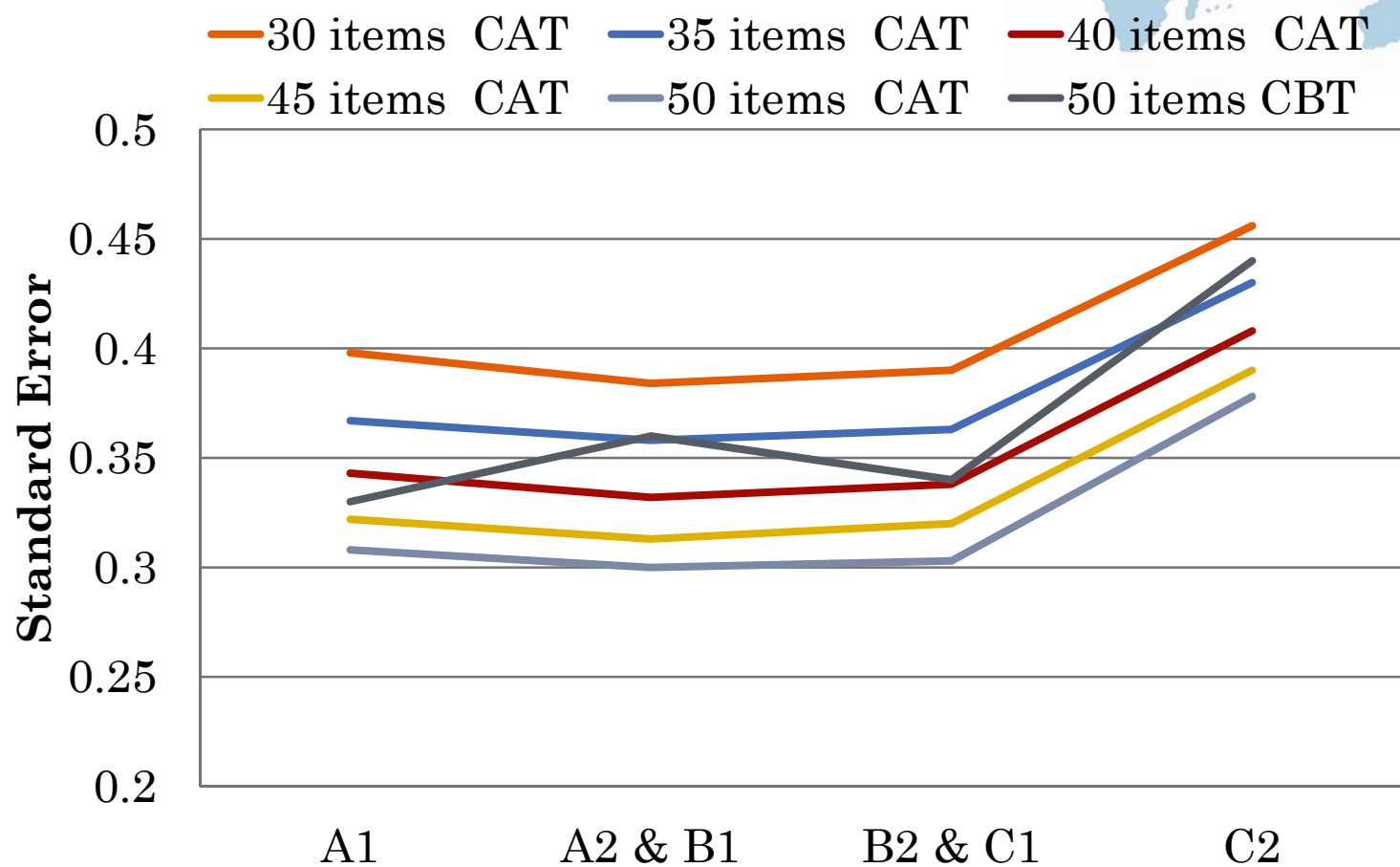
RESULT OF SIMULATION 1

○ listening test



RESULT OF SIMULATION 1

○ Reading test



RESULT OF SIMULATION 2

Listening test

	Selection rules	Generated						
		No Pass	A1	A2	B1	B2	C1	C2
Estimated	No Pass	MI	0.39	0.38				
		MI+topic	0.39	0.38				
		MI+topic+type	0.39	0.38				
		MI+topic+type+items	0.39	0.38				
	A1	MI	0.38	0.39	0.36	0.36		
		MI+topic	0.38	0.39	0.36	0.36		
		MI+topic+type	0.38	0.38	0.36	0.36		
		MI+topic+type+items	0.38	0.39	0.36	0.36		
	A2	MI		0.36	0.36	0.36		
		MI+topic		0.36	0.35	0.35		
		MI+topic+type		0.36	0.35	0.36		
		MI+topic+type+items		0.36	0.36	0.36		
	B1	MI		0.36	0.35	0.36	0.36	
		MI+topic		0.35	0.35	0.36	0.36	
		MI+topic+type		0.35	0.36	0.36	0.36	
		MI+topic+type+items		0.35	0.35	0.36	0.36	
	B2	MI			0.37	0.37	0.38	
		MI+topic			0.36	0.37	0.37	
		MI+topic+type			0.37	0.37	0.38	
		MI+topic+type+items			0.37	0.37	0.38	
	C1	MI			0.38	0.39	0.40	0.42
		MI+topic				0.39	0.40	0.41
		MI+topic+type				0.39	0.40	0.42
		MI+topic+type+items				0.39	0.40	0.42
	C2	MI					0.45	0.48
		MI+topic					0.45	0.46

RESULT OF SIMULATION 2

Reading test

			Generated						
		Selection rules	No Pass	A1	A2	B1	B2	C1	C2
Estimated	No Pass	MI	0.37	0.37					
		MI+topic	0.38	0.38					
		MI+topic+type	0.38	0.38					
		MI+topic+type+items(1)	0.38	0.38					
	A1	MI	0.37	0.36	0.36	0.36			
		MI+topic	0.38	0.37	0.37	0.37			
		MI+topic+type	0.38	0.37	0.37				
		MI+topic+type+items(1)	0.39	0.38	0.37	0.38			
	A2	MI		0.36	0.36	0.35			
		MI+topic		0.36	0.36	0.36			
		MI+topic+type		0.37	0.37	0.37			
		MI+topic+type+items(1)		0.37	0.37	0.37			
	B1	MI			0.35	0.35	0.36	0.36	
		MI+topic		0.36	0.36	0.36	0.36	0.36	
		MI+topic+type		0.36	0.36	0.36	0.36	0.37	
		MI+topic+type+items(1)		0.37	0.36	0.36	0.36	0.36	
	B2	MI				0.36	0.36	0.36	
		MI+topic				0.36	0.37	0.37	
		MI+topic+type				0.37	0.37	0.38	
		MI+topic+type+items(1)				0.37	0.37	0.38	
	C1	MI				0.37	0.37	0.39	0.46
		MI+topic				0.38	0.39	0.42	0.49
		MI+topic+type				0.38	0.39	0.42	0.50
		MI+topic+type+items(1)				0.38	0.39	0.41	0.47
		MI							
		MI+topic							

CONCLUSION

○ Simulation Study 1

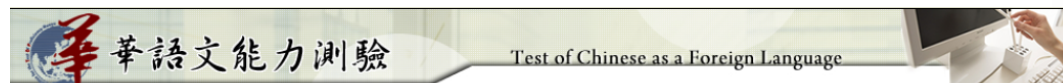
- In CAT, the standard errors for all passing score were lower than that of non-adaptive test (CBT) with the same test length. It can save about 30% of items.
- More difficult items need to be included in the CAT items bank to enhance the efficiency for examinees with higher proficiency level.
- When using MI item selection method, the measurement precision was decreased when the number of test length increased.

○ Simulation Study 2

- The “MI+topic+type” item selection rules showed the lowest differences compared with MI. The “items(testlet)” constraints showed the highest standard errors as compared with MI procedure, especially for the high proficiency levels.
- The percentage of correctly classify (PCC) were 54%~83% for listening test and 60% ~85% for reading test. The extended PCC (± 1 level) were about 100%.



TOCFL CAT 適性化測驗



2)



題數：2/22 精準度：0.3

題目編號：E00125900

總題

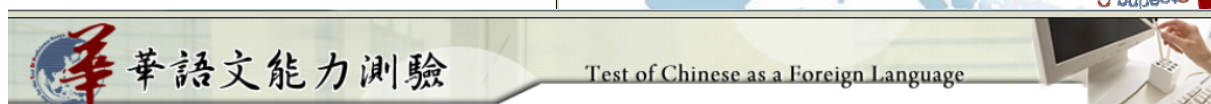
估計誤差：2

- ☐ (A) 她請這位先生趕快上車
- ☒ (B) 十點的車就是最後一班車
- ☐ (C) 等一會兒就沒有車可以坐了
- ☐ (D) 她請這位先生等十點半的車

D



Steering Committee for th



59 : 19

10)

許多開發中的國家，為了改善國家的經濟狀況，利用農業或天然資源，開始高度發展工業或商業。隨之而來的經濟發展，創造大量的工作機會，人民的收入提高了，經濟也就變好了。另一方面也引發了新的問題，城市和鄉村的人口不平均、環境污染等，這些代價都是開發中國家的政府所必須面對的。

題數：10/22 精準度：0.3

題目編號：F00065210

總題數：30 難度：0.111

估計誤差：0.907 類型：短文閱讀

能力值：0.000

題組題編號：F00065200

本文主要在談論什麼主題？

- ☐ (A) 發展經濟的新作法
- ☐ (B) 發展農業後的影響
- ☐ (C) 改善經濟所需的資源
- ☐ (D) 經濟發展的相關問題

D



Steering Committee for the Test Of Proficiency - Huayu

Thank You !



chenph@ntnu.edu.tw