

# **Powerful Tests, Powerless Test Designers?**

**— New Challenges Facing the  
College English Test**

# Testing in China

## ▶ Imperial Examination system

## ▶ Testing (考: **Kao**) in present-day China

- 中考 (Zhōng Kǎo): high school entrance
- 高考 (Gāo Kǎo): college/university entrance
- 考研 (Kǎo Yán): graduate program entrance
- 考博 (Kǎo Bó): doctoral program entrance

.....

- TEM, CET, PETS ...
- TOEFL, IELTS, TOEIC, BEC, FCE, BULLETS ...

# The power of tests

Tests have a very powerful role in modern societies.

Shohamy 2001: 113

Their results have detrimental effects on individuals as they can create

- winners and losers,
- successes and failures,
- rejections and acceptances.

G o Kǎo: college/university entrance

# The power of tests

- ▶ **levers**  
of instructional reform  
(Chapman and Snyder 2000)
- ▶ **lever**  
for change  
(Pearson 1988: 101)
- ▶ **catalyst**  
for curricular innovation
- ▶ **strategy**  
to promote curricular innovation
- ▶ **mechanism**  
to drive instruction

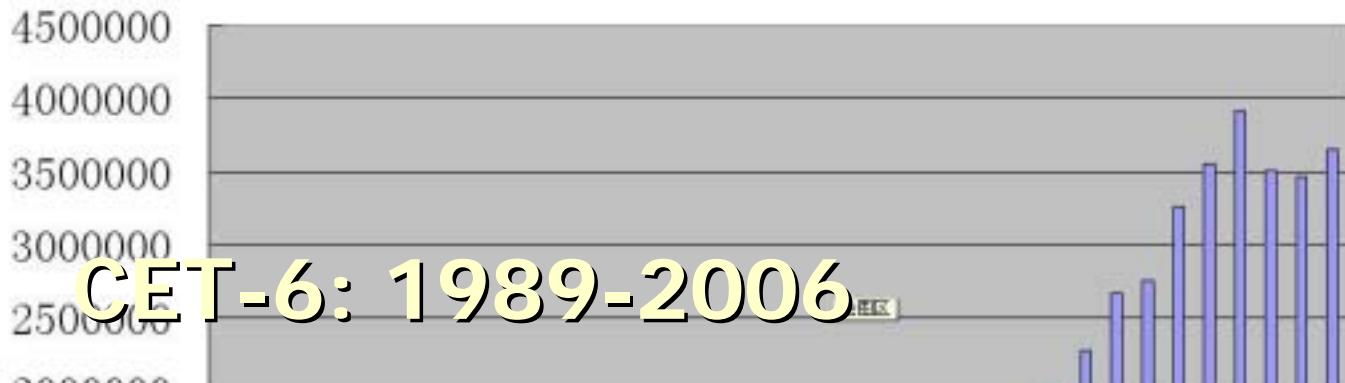
- ▶ **ferocious master**  
of the educational process  
(Madaus 1988: 84-85)
- ▶ **disempowers**  
the great majority of teachers  
(Shepard 1991: 27)
- ▶ **unethical and undemocratic way**  
of making policy  
(Shohamy 2000: 11)

**Andrews 2004: 39-42**

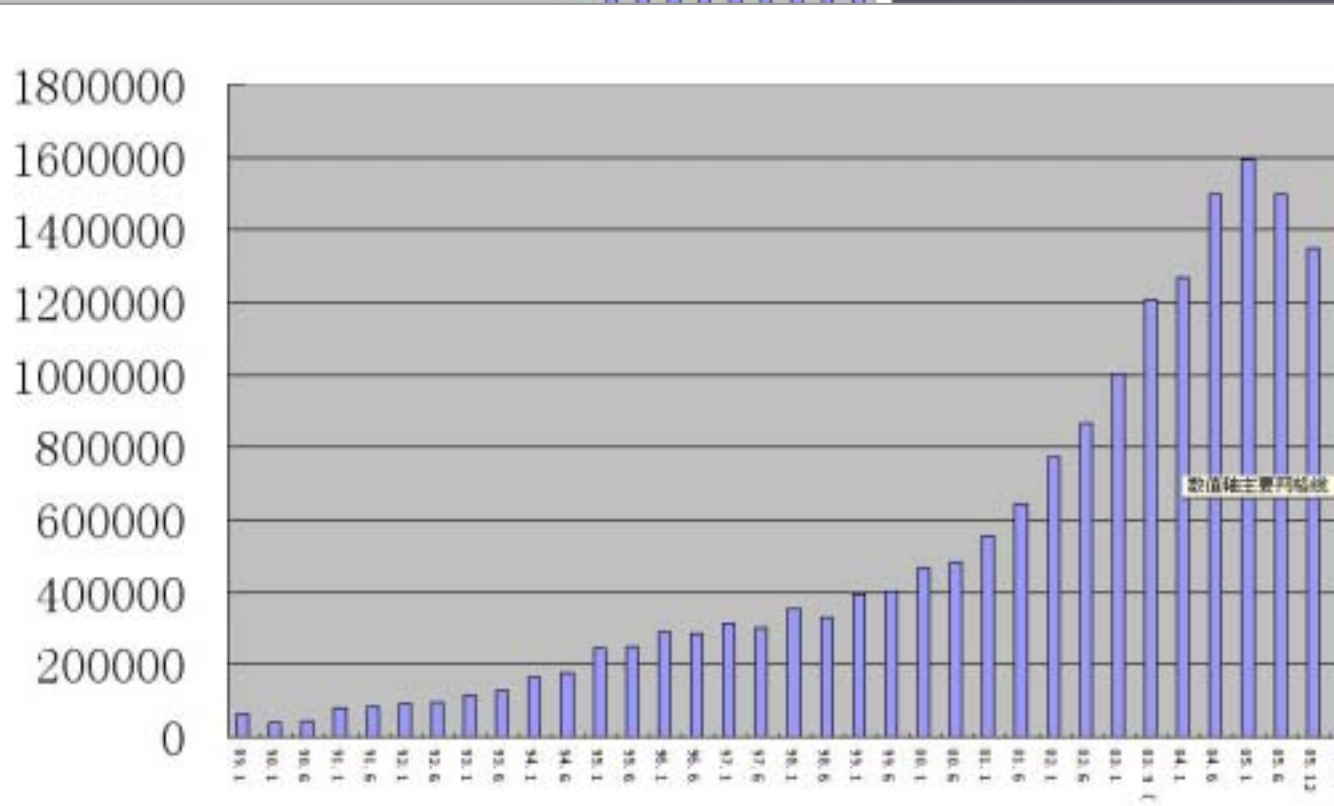
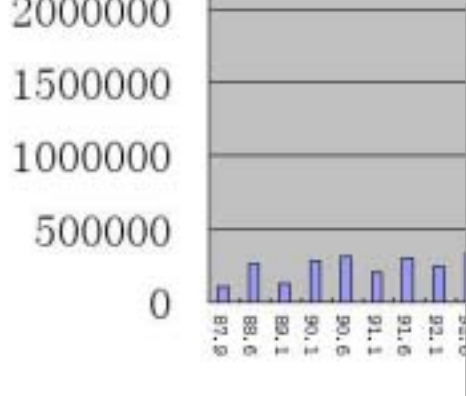
# Large-scale: What does it mean in China?

	2005	2006
Gao Kao	8.67 million	9.50 million
Kao Yan	1.17 million	1.27 million
CET	<b>10.5 million</b>	<b>12 million</b>
PETS	800,000	900,000
TOEFL	120,000	21960
IELTS	100,000	130,000

# CET-4: 1987-2006



# CET-6: 1989-2006



# CET: past, present and future

## Part 1 How powerful is the CET?

- A brief history of the CET

## Part 2 What has been done to tackle the power?

- Recent reform of the CET

## Part 3 What can be done to further improve the washback of the CET?

- The way forward for the CET

# Questions to be focused on

**How powerful can a test be?**

**Do test designers have any power over the use of tests?**

**What are the responsibilities of test designers?**

# 1. CET: A brief history

## Test purposes

... tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of society at large.

Bachman (1990: 279)

# The intended purposes of the CET

A pressing need for university graduates with a higher level of English language proficiency

## College English Teaching Syllabus



# College English Teaching Syllabus

- 1985/1986:

Ability to read in English; Use English as a medium to access information related to their major areas of study

- 1999:

Reading + listening, speaking and writing abilities

- 2004:

Further develop listening and speaking abilities;  
Exchange information effectively through both spoken and written channels

# The intended purposes of the CET

- ▶ Measure the English language proficiency of college and university students
- ▶ Promote the teaching and learning of English as a foreign language in colleges and universities in China (the implementation of the NCETS)

# Beneficial effects

```
graph TD; A[Beneficial effects] --> B[CE teaching]; A --> C[CE learning]; A --> D[Society];
```

## CE teaching

- ✓ more attention
- ✓ more resources
- ✓ more effective teaching:  
feedback  
information for  
improving teaching

## CE learning

- ✓ more attention
- ✓ more resources
- ✓ more effective learning:  
feedback  
information for  
improving learning

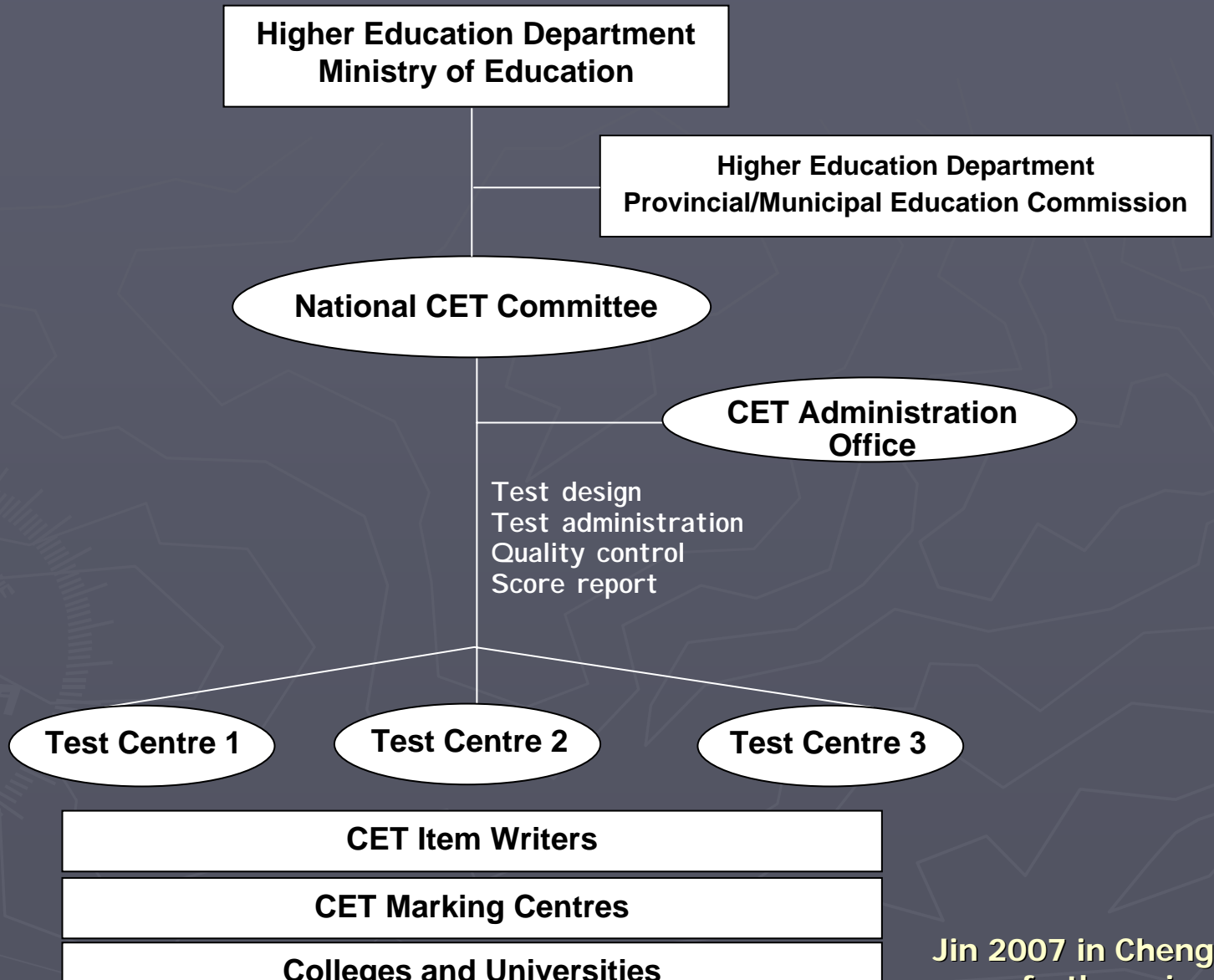
## Society

- ✓ objective  
evaluation of a  
student's overall  
English  
proficiency
- ✓ information for  
making  
educational  
policies

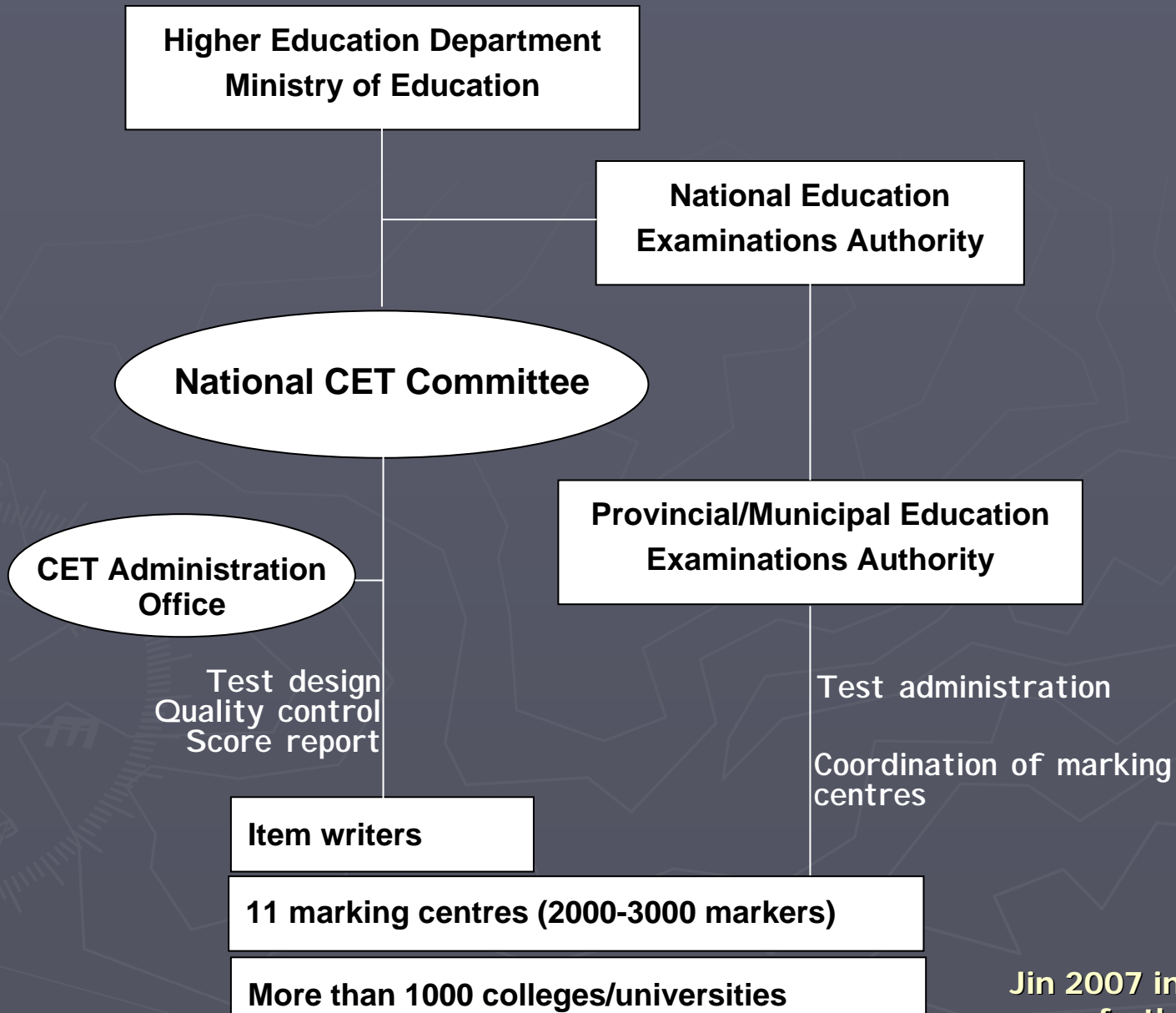
# The National CET Committee

- ❑ An academic organization
- ❑ CET Design Group (1986-1993)
  - ▶ 12 language teaching and testing experts
- ❑ National CET Committee (1994-now)
  - ▶ 3<sup>rd</sup> NCETC: 25 professors from 23 universities
  - ▶ leading professors at the forefront of applied linguistics research in China

# Operational structure of the CET (1987-2005)



# Operational structure of the CET (since 2006)



# Efforts to maintain the professional standards of the CET

- ▶ **Test validity:** CET validation study
  - Componential view of the language construct
  - Content representativeness and coverage
  - Concurrent validity, predictive validity, face validity...
- ▶ **Test reliability:** A strict system of quality control
  - intra-marker consistency
  - inter-marker consistency
  - inter-centre consistency

# Technical data (CET-4 Prototype)

	For the whole objective part of 90 items
KR-20	0.93
SEM	4.56

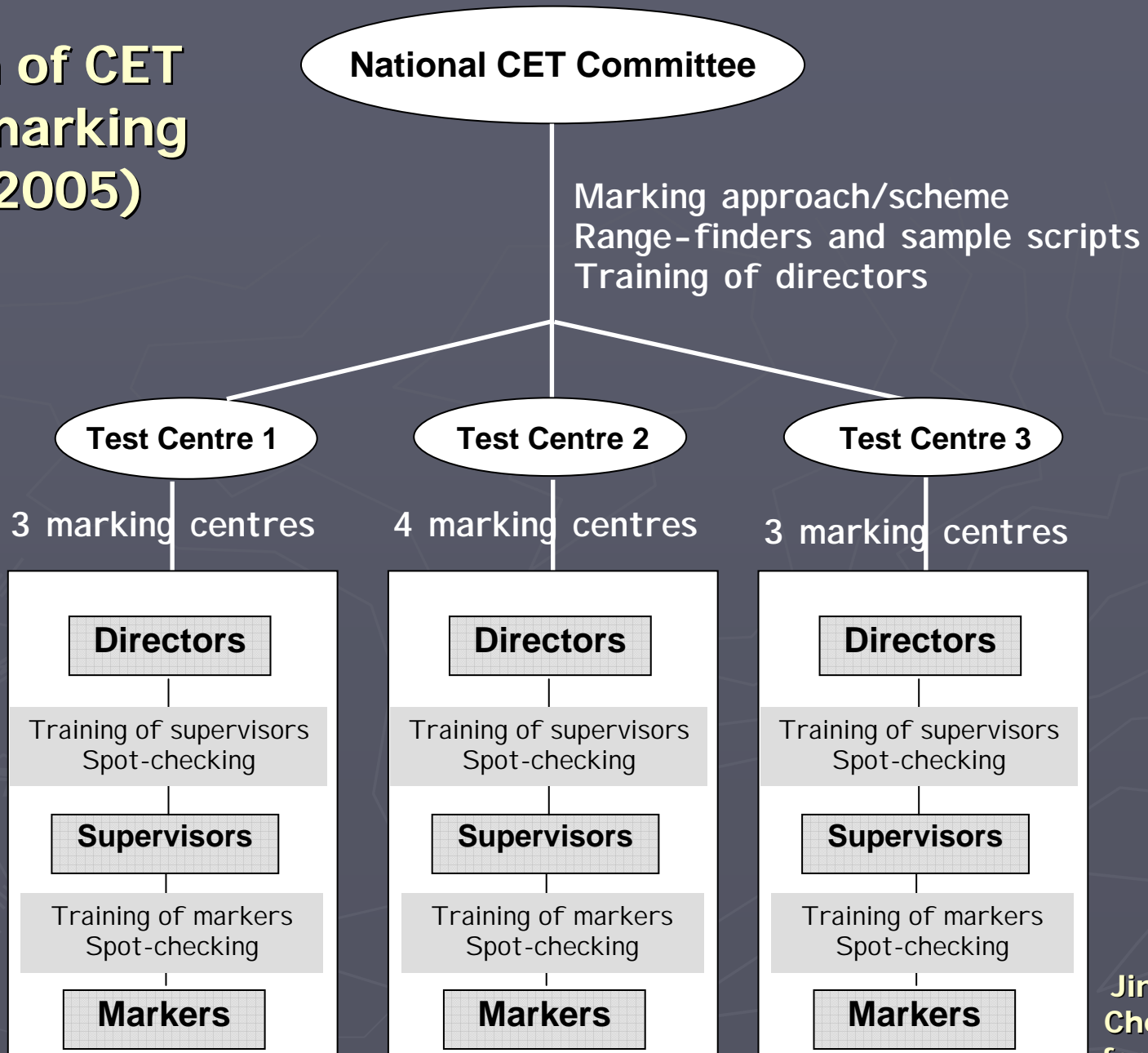
Difficulty value	< 0.1	0.1 ~ 0.2	0.2 ~ 0.3	0.3 ~ 0.4	0.4 ~ 0.5	0.5 ~ 0.6	0.6 ~ 0.7	0.7 ~ 0.8	0.8 ~ 0.9	0.9 ~ 1.0	Mean = 0.653
No. of items				1	10	19	25	25	8	2	

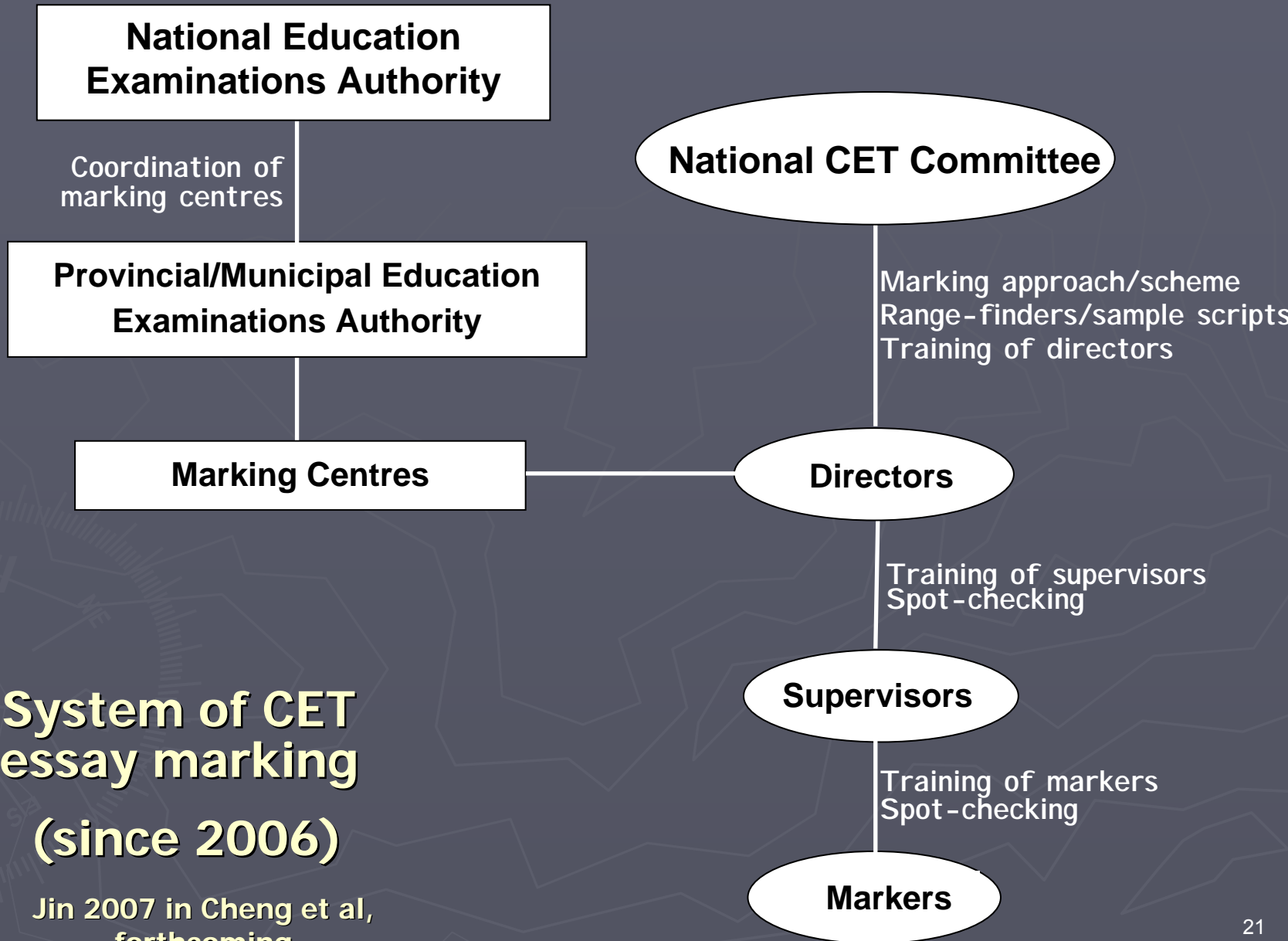
Reliability value	< 0.1	0.1 ~ 0.2	0.2 ~ 0.3	0.3 ~ 0.4	0.4 ~ 0.5	0.5 ~ 0.6	0.6 ~ 0.7	0.7 ~ 0.8	0.8 ~ 0.9	0.9 ~ 1.0	Mean = 0.430
No. of items		3	10	17	31	26	3				

## Technical data (CET-4 Prototype)

	Listening	Reading	Voc & Str	Cloze	Writing
Listening	1.000				
Reading	0.563	1.000			
Voc & Str	0.539	0.615	1.000		
Cloze	0.467	0.531	0.626	1.000	
Writing	0.387	0.359	0.470	0.404	1.000
Total	0.792	0.892	0.802	0.708	0.581

# System of CET essay marking (1987-2005)





**System of CET  
essay marking  
(since 2006)**

Jin 2007 in Cheng et al,  
forthcoming

CET conference marking

CET online marking



## Remarks by the Vice Minister of Education

... The fact that such a large-scale test has been developing steadily in the past 17 years was in itself solid evidence to show that the CET has met the social needs, won social recognition, produced beneficial effects on society and contributed significantly to the continued improvement of the quality of College English teaching in China.

# The National CET Committee

- ▶ Exchanges with UCLES, ETS, HKEAA, CEEC, LTTC, STEP, KELTA, etc.
- ▶ AFELTA: Academic Forum on English Language Testing in Asia
  - Initiator/organizer and host of the 1<sup>st</sup> and 7<sup>th</sup> AFELTA
- ▶ International conferences
  - Symposium at 2000 LTRC, Vancouver
  - 1<sup>st</sup> International Conference on English Language Testing in China, 2002, Shanghai
  - “Big Tests” Symposium, 2005 AILA, Wisconsin

## 2. Reform of the CET: Challenges facing the NCETC

### Uses of the CET for purposes not intended

- Requirement for college/university graduation
- Requirement for Bachelor's degree
- Prerequisite to getting a job
- Prerequisite to getting residential certification

# The power of high-stakes tests

Results of the test “are seen—rightly or wrongly—by students, teachers, administrators, parents, or the general public, as being used to make important decisions that immediately and directly affect them”

Madaus, 1988: 87

# Negative effects

```
graph TD; A[Negative effects] --> B[CE teaching]; A --> C[CE learning]; A --> D[Society];
```

## CE teaching

### X teaching to the test

- use coaching materials
- narrow the teaching content
- stop teaching after students pass the test

## CE learning

### X learning for the test

- use coaching materials
- narrow the learning content
- aim at a narrow pass and stop learning after a narrow pass

## Society

### X inappropriate use of CET certificate

- e.g. when recruiting new employees for jobs that do not require English

# What's the point of a powerful language test ...

if the test cannot have a good and happy effect on  
millions of **EFL learners in China**?

# Reform of the CET: What has been done

## ► Changes over the past twenty years

- 1990: separate paper for WRITING
- 1996: new item forms (CD, SAQ, translation, etc)
- 1997: Average Graded Score
- 1998: a minimum score requirement for CET-4 writing
- 1999: CET-SET
- ...

# Recent reform of the CET: Changes in the score reporting system

Starting from	Mean	Sd	Score report		Score interpretation
1987/1989	72	12	Pass	Distinction	Percentile position in the norm group (students in the six top universities in China)
			60	85	
June 2005	500	70	No pass or fail; Reporting total and profile scores		In the process of setting up a new norm
December 2006					

# Changes in the score reporting system

- ▶ New score reporting system : Why?
  - Encourage more appropriate use of the CET test results
  - Reduce social pressure on the CET
- ▶ New score reporting system: How?
  - No pass or fail
  - Issue a score report instead of a certificate
  - Report both the total and profile scores

# Recent reform of the CET: Changes in the test content and format

- ▶ More emphasis on listening ability
- ▶ More attention to contextualized language use instead of context free knowledge of language
- ▶ A larger proportion of constructed response items

# CET test content

	1987 (CET-4) 1989 (CET-6)	1997	1999	2006
<b>Listening</b>	20% <ul style="list-style-type: none"> <li>• Short conversations</li> <li>• Listening passages</li> </ul>	20% <ul style="list-style-type: none"> <li>• Short conversations</li> <li>• Listening passages or Passage dictation</li> </ul>		35% <ul style="list-style-type: none"> <li>• Short conversations</li> <li>• Long conversations</li> <li>• Listening passages and</li> <li>• Passage dictation</li> </ul>
<b>Reading</b>	40% <ul style="list-style-type: none"> <li>• Careful reading</li> </ul>	40% <ul style="list-style-type: none"> <li>• Careful reading</li> </ul>		35% <ul style="list-style-type: none"> <li>• 25% Careful reading</li> <li>• 10% Skimming and scanning</li> </ul>
<b>Vocabulary &amp; structure</b>	15% <ul style="list-style-type: none"> <li>• 7.5% Vocabulary</li> <li>• 7.5% Structure</li> </ul>	15% <ul style="list-style-type: none"> <li>• CET-4: 10% voc. and 5% Structure</li> <li>• CET-6: 15% Voc.</li> </ul>		----
<b>Writing</b>	15% <ul style="list-style-type: none"> <li>• 120-140 words in 30 minutes</li> </ul>	15% <ul style="list-style-type: none"> <li>• 120-140 words in 30 minutes</li> </ul>		20% <ul style="list-style-type: none"> <li>• 130-150 words in 30 minutes</li> <li>• Translation (Chinese to English)</li> </ul>
<b>Integrated tasks</b>	10% <ul style="list-style-type: none"> <li>• Cloze or</li> <li>• Error identification and correction</li> </ul>	10% <ul style="list-style-type: none"> <li>• Cloze or</li> <li>• Error identification and correction</li> </ul>		10% <ul style="list-style-type: none"> <li>• Cloze or</li> <li>• Error identification and correction</li> </ul>
<b>Speaking (optional)</b>	----		<ul style="list-style-type: none"> <li>• Monolog</li> <li>• Dialogue</li> <li>• Group discussion</li> </ul>	Jin 2007 in Cheng et al. forthcoming

# CET test format

	1987 (CET-4) 1989 (CET-6)	1997	1999	2006
Listening	<ul style="list-style-type: none"> <li>MCQ</li> </ul>	<ul style="list-style-type: none"> <li>MCQ</li> <li>Compound Dictation</li> </ul>		<ul style="list-style-type: none"> <li>MCQ</li> <li>Compound Dictation</li> </ul>
Reading	<ul style="list-style-type: none"> <li>MCQ</li> </ul>	<ul style="list-style-type: none"> <li>MCQ</li> <li>SAQ</li> <li>Translation (English to Chinese)</li> </ul>		<ul style="list-style-type: none"> <li>MCQ</li> <li>SAQ</li> <li>Banked Cloze</li> <li>True or False</li> <li>Sentence completion</li> </ul>
Vocabulary & structure	<ul style="list-style-type: none"> <li>MCQ</li> </ul>	<ul style="list-style-type: none"> <li>MCQ</li> </ul>		-----
Writing	<ul style="list-style-type: none"> <li>Essay writing</li> </ul>	<ul style="list-style-type: none"> <li>Essay writing</li> </ul>		<ul style="list-style-type: none"> <li>Essay writing</li> <li>Translation (Chinese to English)</li> </ul>
Integrated tasks	<ul style="list-style-type: none"> <li>Cloze</li> <li>Error identification and correction</li> </ul>	<ul style="list-style-type: none"> <li>Cloze</li> <li>Error identification and correction</li> </ul>		<ul style="list-style-type: none"> <li>Cloze</li> <li>Error identification and correction</li> </ul>
Speaking (optional)	-----		<ul style="list-style-type: none"> <li>Q&amp;A</li> <li>Individual presentation</li> <li>Group discussion</li> </ul>	

# 3. The way forward

**From academic endeavour to social ethics:  
the NCETC's new mission statement**

**... it has already been established that simply changing test contents and methods will NOT necessarily bring about direct and desirable changes in education as intended through a testing change.**

Alderson 2004

# CET: The way forward

- ▶ Intrinsic power of a test
- ▶ Extrinsic power of a test

Li 2000

**Strengthening the intrinsic power of the test**

**Taking and sharing the responsibility for the consequences of the test**

# Strengthening the intrinsic power

## Why?

- ▶ Eliminate potential sources of negative washback
  - construct under-representation
  - construct-irrelevant variance (Messick 1996:252)

## How?

- ▶ Maintain the highest possible quality of the test
- ▶ Develop a common level framework for Chinese EFL learners
- ▶ Explore the possibility of a criterion-referenced CET
- ▶ Develop Computer- or Internet-based CET
- ▶ ...

# Dilemmas facing test designers

... negative effects can result from a properly developed, correctly targeted test that is being implemented in line with all available knowledge about best practice ...

Hamp-Lyons 1997:297

... It should not be assumed that a 'good' test will automatically produce good effects in the classroom, or that a 'bad' test will necessarily produce negative ones.

Dianne Wall 2000:505-506

**the more likely the test is to be misused**



**the higher the stakes**



**the more power the test has**



**the higher its reputation**



**the better the quality of a test**

There are limits to what a test  
developer can achieve ...

JC Alderson in Cheng et al. 2004: foreword

- ▶ Is the test developer responsible for uses and misuses of tests?
- ▶ What is the role of the tester once s/he notices misuses?
- ▶ Or, perhaps it is not the responsibility of the tester at all to worry about the test takers after a test had been handed to the users.

E Shohamy 2001: 115

# What is washback?

1993 J.C. Alderson and Dianne Wall

**Does washback exist?**

2004 J.C. Alderson in Cheng et al (Foreword)

**What does washback look like?**

**What brings washback about?**

**Why does washback exist?**

# What is washback?

**Messick 1989, 1996** → **Consequential validity**

**Bachman and Palmer 1996**  
test usefulness framework → **Impact**

**McNamara 2000** → **Impact and washback**

**Kunnan 2004**  
test fairness framework → **Social consequences**

There was a growing awareness of the importance of factors other than test design in determining whether tests would have the impact that was desired.

Dianne Wall 2000: 499

# Tests used as power tools

Tests, therefore, are not neutral but rather embedded in **political, social, educational, ideological and economic** contexts.

... the **powerful uses of tests in society as tools** for mostly political, social, educational and ideological agendas

E Shohamy 2001: 113-114

# Taking and sharing the responsibility

1. Investigate the social consequences
  - ❑ Identify the CET stake-holders/factors involved
  - ❑ Investigate the washback on teaching/learning and the impact on society (In which direction? In what way?)
2. Make other users aware
  - ❑ Disseminate relevant information
  - ❑ Introduce better practices (e.g. performance assessment)
3. Share responsibilities
  - ❑ Engage all the stake-holders in the mission:  
Every stake-holder has a role to play!

# Code of Ethics for ILTA

1. Respect for test takers
2. Confidence
3. Research
4. No professional misuse
5. Continue to develop professionally
6. Uphold profession
7. Improve professional quality
8. Obligations to own society
9. **Consequential effects**

Code of Ethics for ILTA. 2000. Adopted at the annual meeting of ILTA held in Vancouver, March 2000

“Language testers shall regularly consider the potential effects, both short and long term, on all stakeholders of their projects, reserving the right to withhold their professional services on the grounds of conscience.”

# The future of the CET

... lies in its role as an impetus to English language teaching and learning at the tertiary level in China. The job of the CET Committee is not simply of an academic nature ...

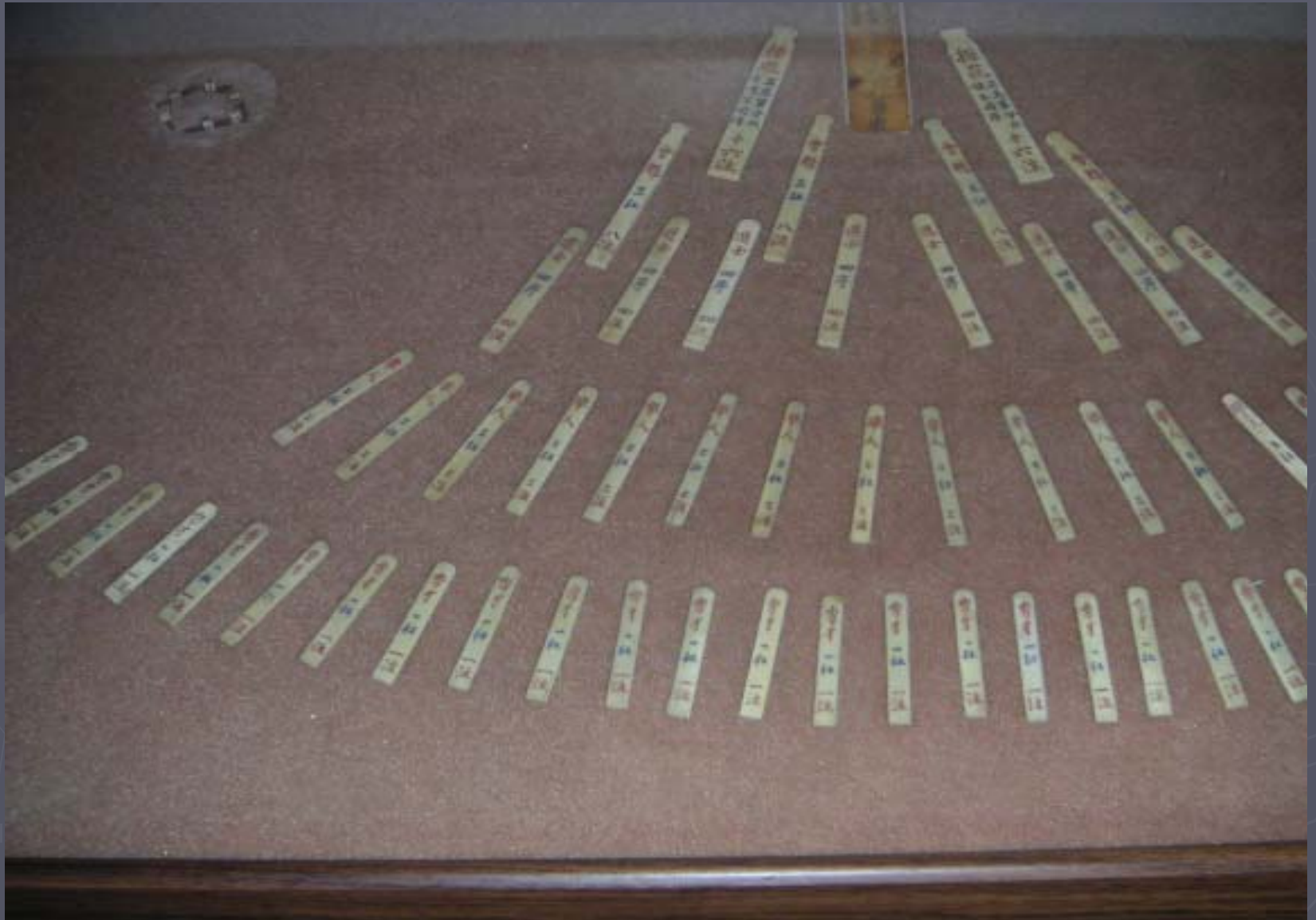
The widening of the lens through which we look at tests brings with it great **ethical or moral** problems about the responsibility of the language tester, and of language testing as a profession.

Hamp-Lyons 2000: 587

Thank you for your attention!

[yjin@sjtu.edu.cn](mailto:yjin@sjtu.edu.cn)

# Levels of examinations in the Imperial system



# Penalties for cheating



严惩枪手 清代

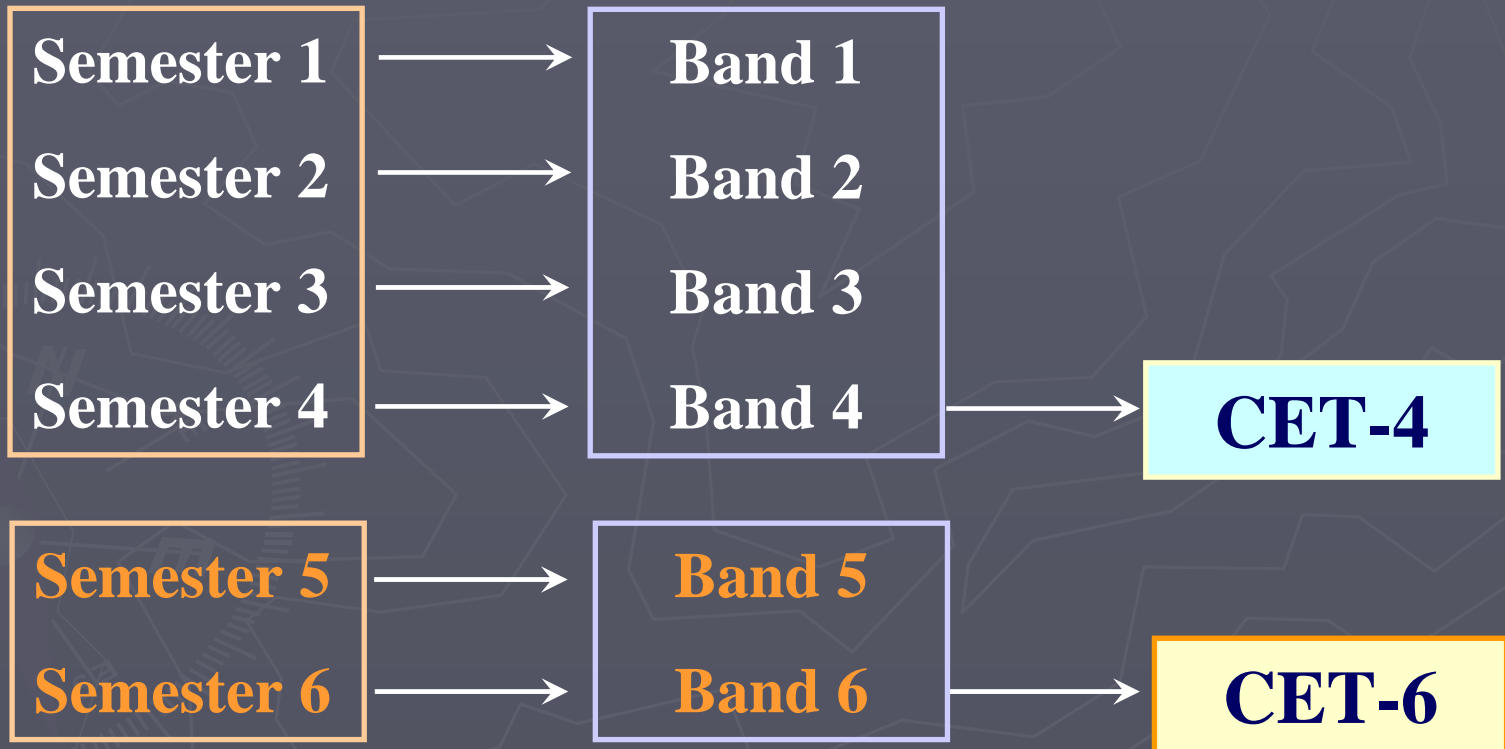
代笔者称“枪手”，考试时替人代笔称“枪手”。图为枪手被查出后  
枷号于贡院大门前示众问罪。

# CET-4 and CET-6

**Semester**

**College English Course**

**College English Test**



**Back!**

# CET stakeholders: Who are they?

- ▶ College English teachers
- ▶ University students

## At the micro levels

teaching, learning,  
curriculum and material,  
test preparation and  
attitude, etc.

- ▶ Parents
- ▶ The general public
- ▶ Employers
- ▶ School administrators
- ▶ Educational policy makers

## At the macro levels

educational context,  
social/political/economic  
factors

# Questions to be addressed

## EFL teaching

- ❑ In what way and to what extent is college English teaching affected by the reform of the CET?

## EFL learning

- ❑ In what way and to what extent is college English learning affected by the reform of the CET?

## Educational policies

- ❑ In what way and to what extent are educational policies affected by the reform of the CET?

## Other users

- ❑ In what way and to what extent are other users of the test affected by the reform of the CET?