Cambridge English

NAFLE, Beijing 2011

The Implementation of a VRIP-Q framework for language testing

Nigel Pike
Director Assessment
University of Cambridge ESOL Examinations

Cambridge English

A framework for

	Validity
	Reliability
Positive	**I**mpact
	Practicality
Within a	**Q**uality Management framework

in large-scale international language testing

Cambridge English

The context

Evidence of language ability is being used increasingly as one of the key criteria for life-changing decisions, such as immigration, education and employment.

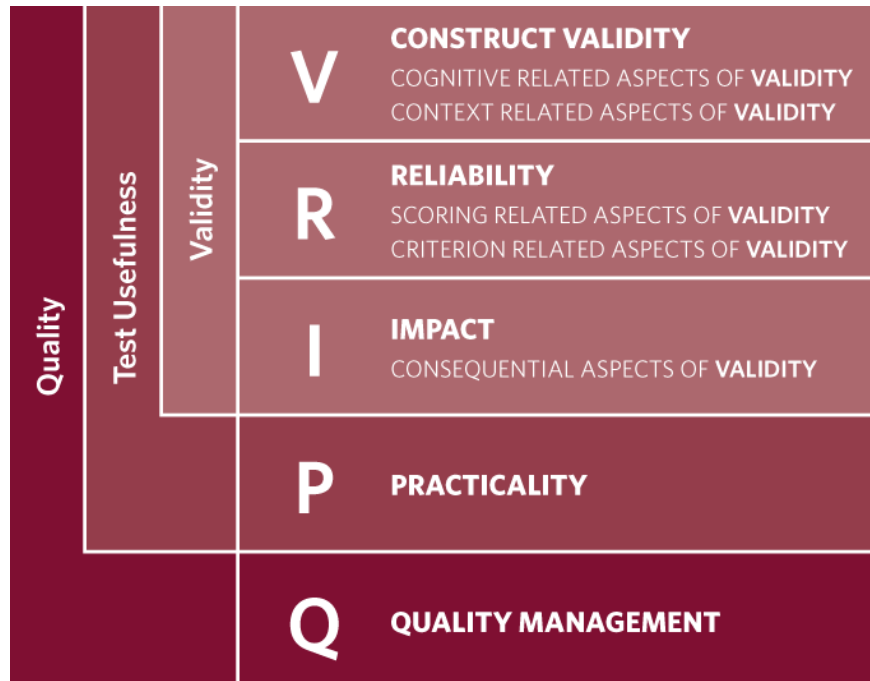**= High stakes and high responsibility**

Cambridge English

International language tests need to be:

- fair

- accurate

- valid

- underpinned by research

- underpinned by robust organisational procedures

Cambridge English
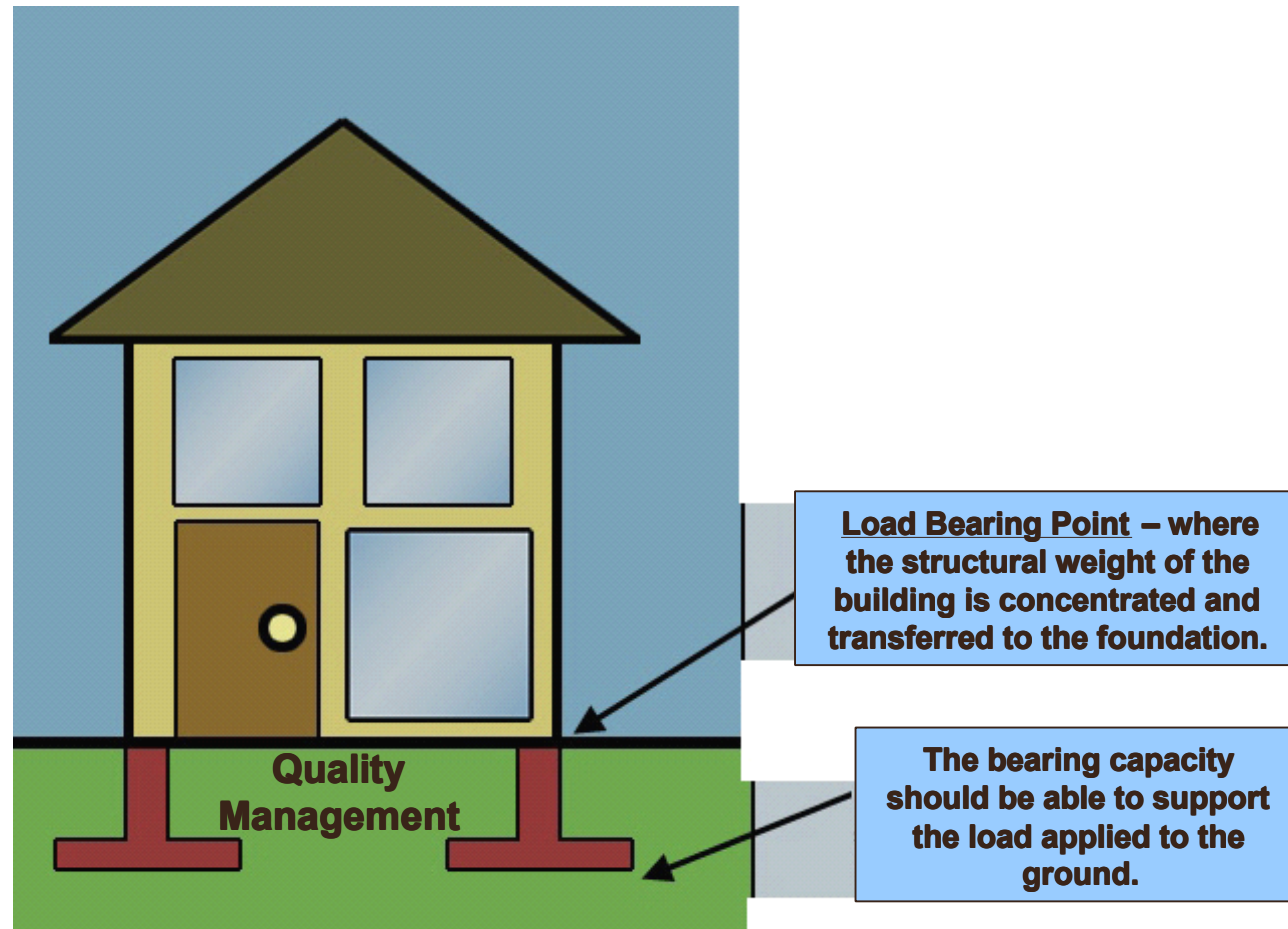
NAFLE, Beijing 2011

The development of a **VRIPQ** framework

Quality Management

Load Bearing Point – where the structural weight of the building is concentrated and transferred to the foundation.

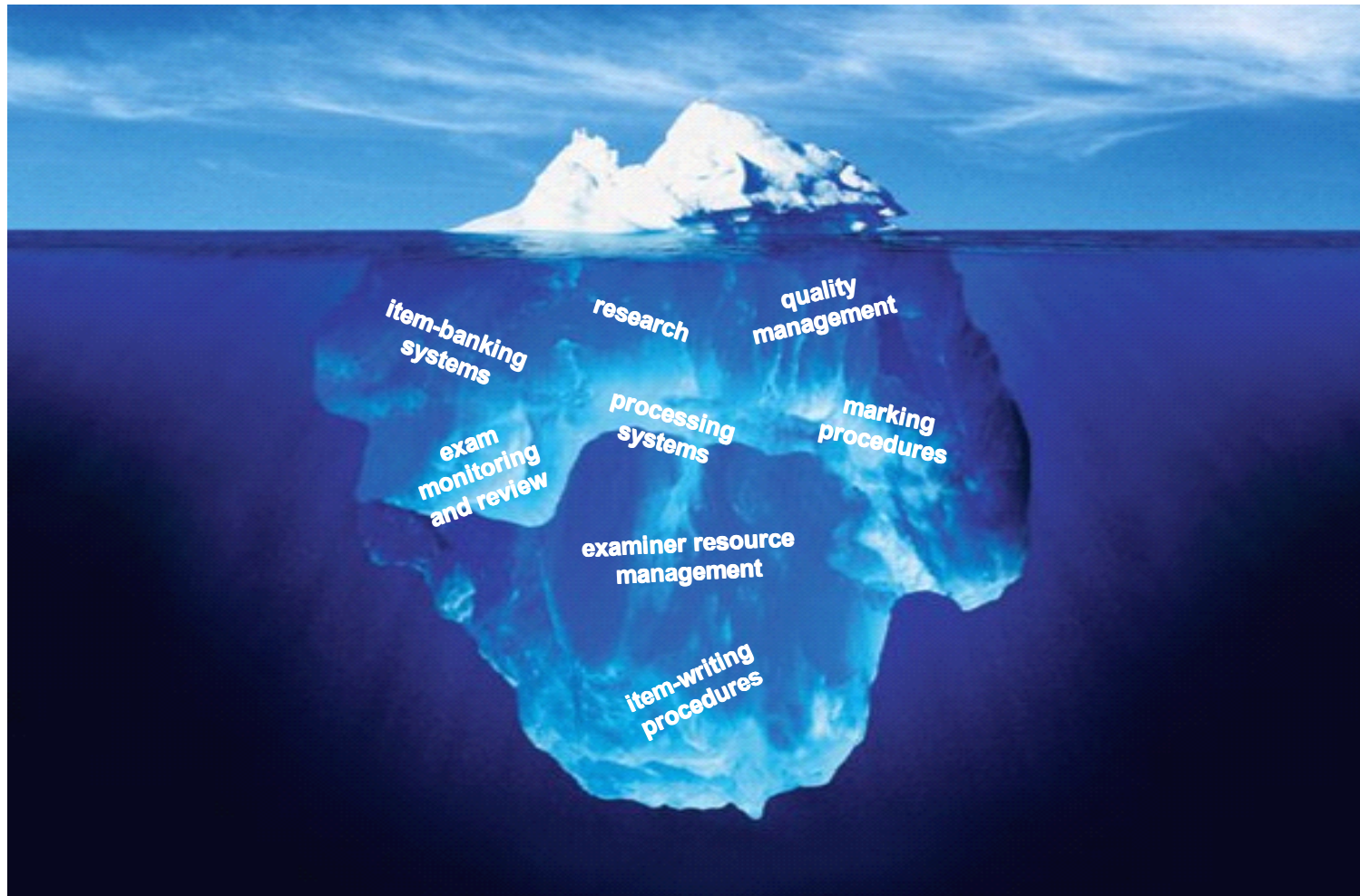The bearing capacity should be able to support the load applied to the ground.

Cambridge English

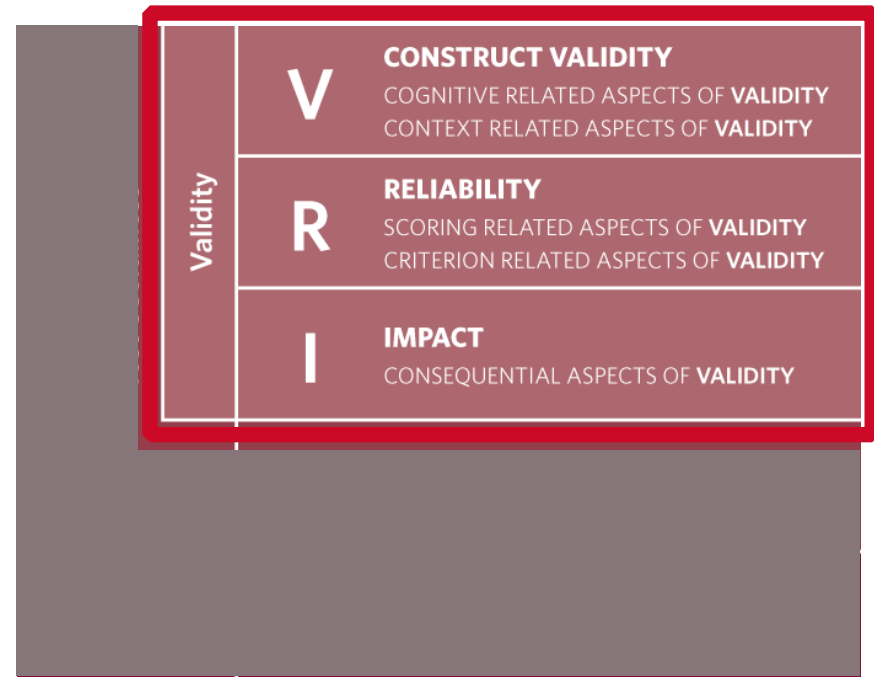## NAFLE, Beijing 2011

Cambridge English

Validity

A test is valid if it measures what we intend it to measure

OR

'The ongoing process of demonstrating that a particular interpretation of test scores is justified.' (Bachman and Palmer, 1996)

Validity	V	**CONSTRUCT VALIDITY** COGNITIVE RELATED ASPECTS OF **VALIDITY** CONTEXT RELATED ASPECTS OF **VALIDITY**
	R	**RELIABILITY** SCORING RELATED ASPECTS OF **VALIDITY** CRITERION RELATED ASPECTS OF **VALIDITY**
	I	**IMPACT** CONSEQUENTIAL ASPECTS OF **VALIDITY**

Validity

Construct validity

The theory of language ability the test is based on

- Models of communicative language ability (Canale and Swain, 1988; Bachman 1990)

- Socio-cognitive model (Weir, 2005)

- Common European Framework of Reference (CEFR – Council of Europe)

Cambridge ESOL approach

Language proficiency in terms of language users ' overall communicative ability sub-divided into skills and sub-skills

⇨ a profile of communicative language ability for an individual candidate

Validity

Cognitive- and context-related aspects of validity

- **cognitive-related validity**

  the extent to which cognitive processes employed by the candidate are the same as those needed in real-world contexts (Target Language Use – TLU)

Cambridge English

Validity

Cognitive- and context-related aspects of validity

- **context-related validity**

  the conditions under which the test is performed e.g. the tasks, rubrics and test administration conditions

  appropriateness of linguistic and content demands

  good sampling of the syllabus

Validity

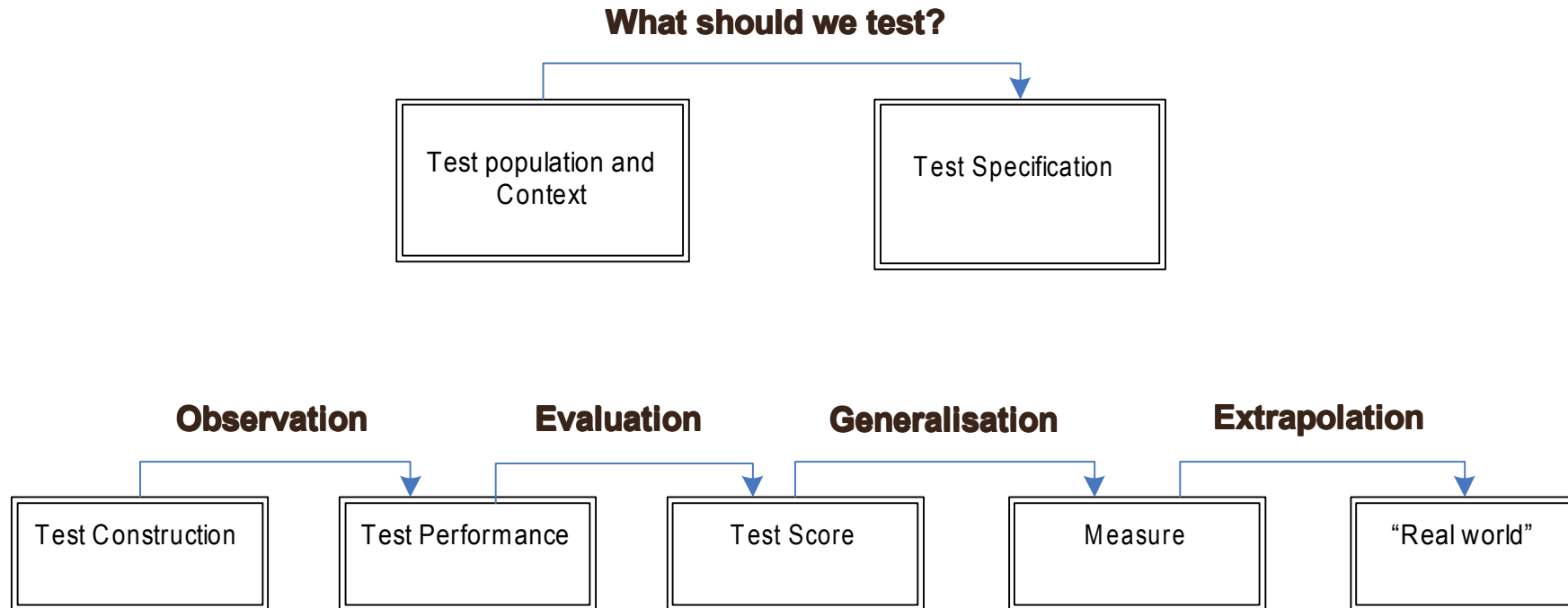**Cognitive- and context-related aspects of validity**

Validation of these aspects should include analysing the degree to which the sample of items and tasks in an exam represent TLU domain.

A variety of tasks and response types can test a wider range of underlying abilities and contribute to an exam's fairness.

Tests need to demonstrate a level of authenticity in both contextual and cognitive terms in order for the generalised inferences to be valid.

Cambridge English

Validity within a test development cycle

What should we test?

Test population and Context	Test Specification

Observation — **Evaluation** — **Generalisation** — **Extrapolation**

Test Construction	Test Performance	Test Score	Measure	"Real world"

Cambridge English

Validity

V	**CONSTRUCT VALIDITY**	COGNITIVE RELATED ASPECTS OF **VALIDITY** CONTEXT RELATED ASPECTS OF **VALIDITY**
R	**RELIABILITY**	SCORING RELATED ASPECTS OF **VALIDITY** CRITERION RELATED ASPECTS OF **VALIDITY**
I	**IMPACT**	CONSEQUENTIAL ASPECTS OF **VALIDITY**

NAFLE, Beijing 2011

Reliability

The extent to which test results are stable, consistent and free from errors of measurement.

Day

Test taker

Error

Marker

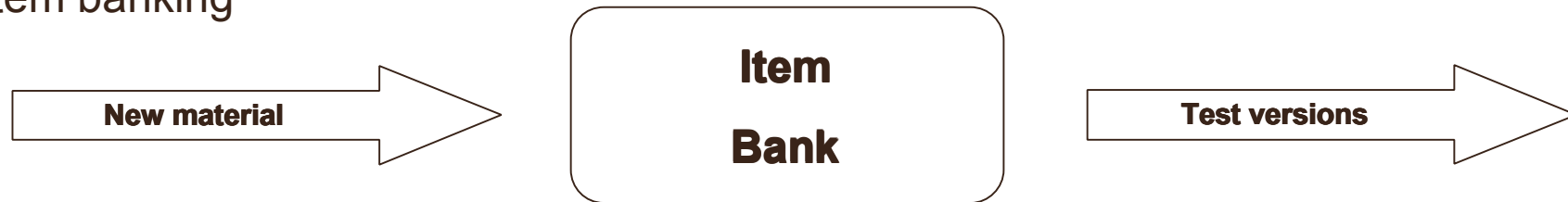Test version

Other
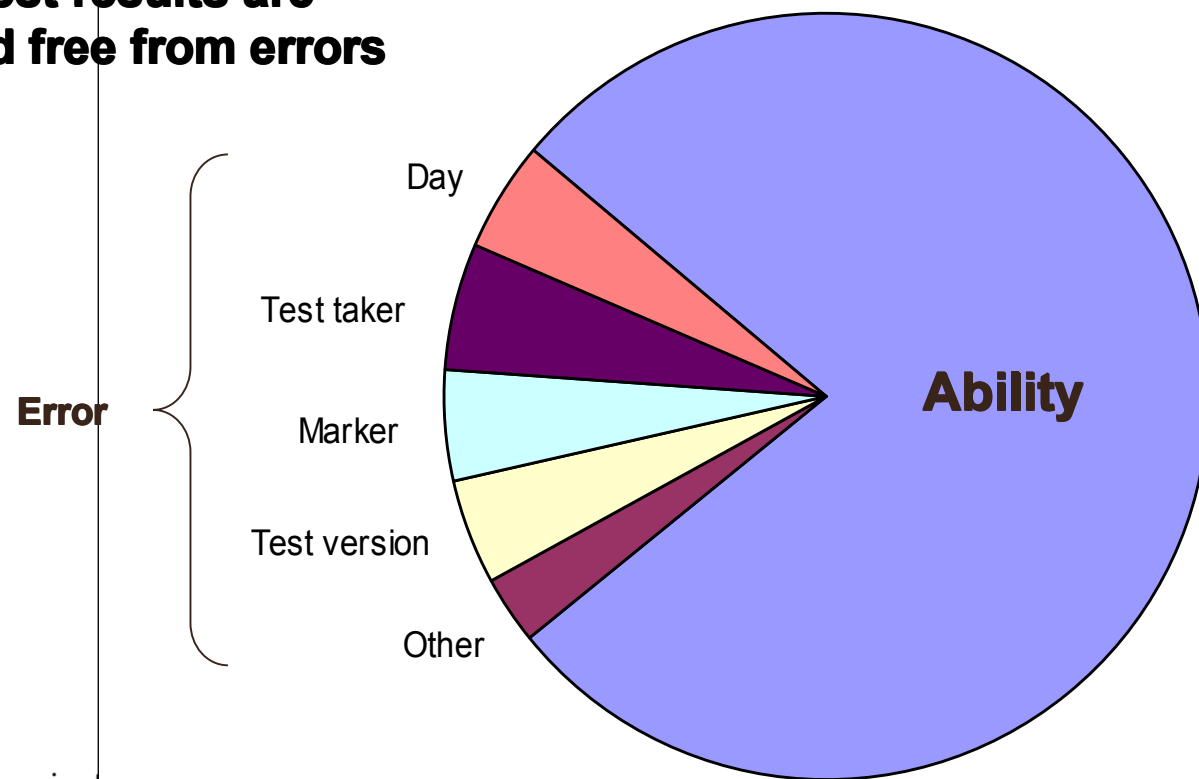
Ability

Reliability

Criterion-related aspects of validity

- criterion-referenced tests are ones in which candidates are assessed against specific target behaviors (e.g. the CEFR)

- setting and maintaining of standards – different test versions must be comparable and so the same underlying ability is needed to 'pass' each test version.

   ▪ item-banking and common scales

Cambridge English

Reliability

Item banking

```
New material  ⟶   Item Bank   ⟶  Test versions
```

- rigorous quality assurance

- calibration (Rasch analysis) to a single measurement scale

⇨ can construct tests of known difficulty and ensure equivalence across test versions and standardised results

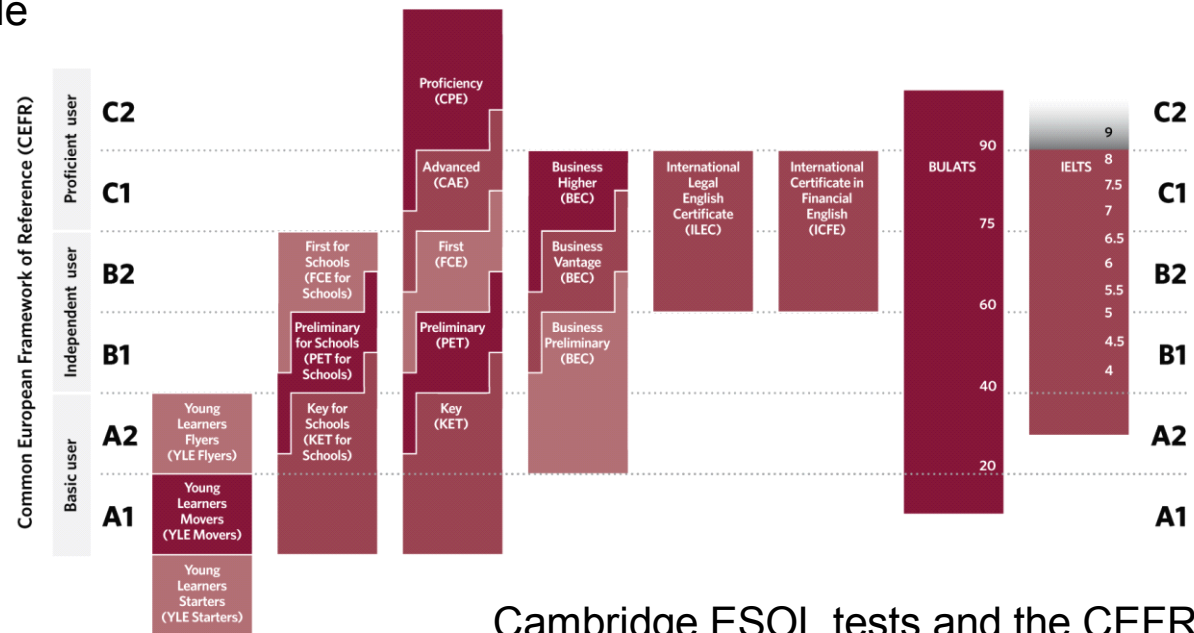- common scales for Writing and Speaking against stated criteria

Cambridge English

Reliability

Criterion-related aspects of validity (continued)

Use of scales: an example

Cambridge ESOL tests and the CEFR

Reliability

Scoring-related aspects of validity

Key aim is to minimise effects of sources of error

e.g. standardising marking and test administration conditions

Standardising marking

• worldwide team leader system

• standardisation of marking

      • online capture of marks – training and live

      • electronic marking of Writing and 'seeded scripts'

Cambridge English

Validity

V	**CONSTRUCT VALIDITY** COGNITIVE RELATED ASPECTS OF **VALIDITY** CONTEXT RELATED ASPECTS OF **VALIDITY**
R	**RELIABILITY** SCORING RELATED ASPECTS OF **VALIDITY** CRITERION RELATED ASPECTS OF **VALIDITY**
I	**IMPACT** CONSEQUENTIAL ASPECTS OF **VALIDITY**

Cambridge English

Impact

Consequential aspects of validity

Assessment can have important effects and consequences

- within a classroom (washback)

- within an education system

- within society more widely

Cambridge English

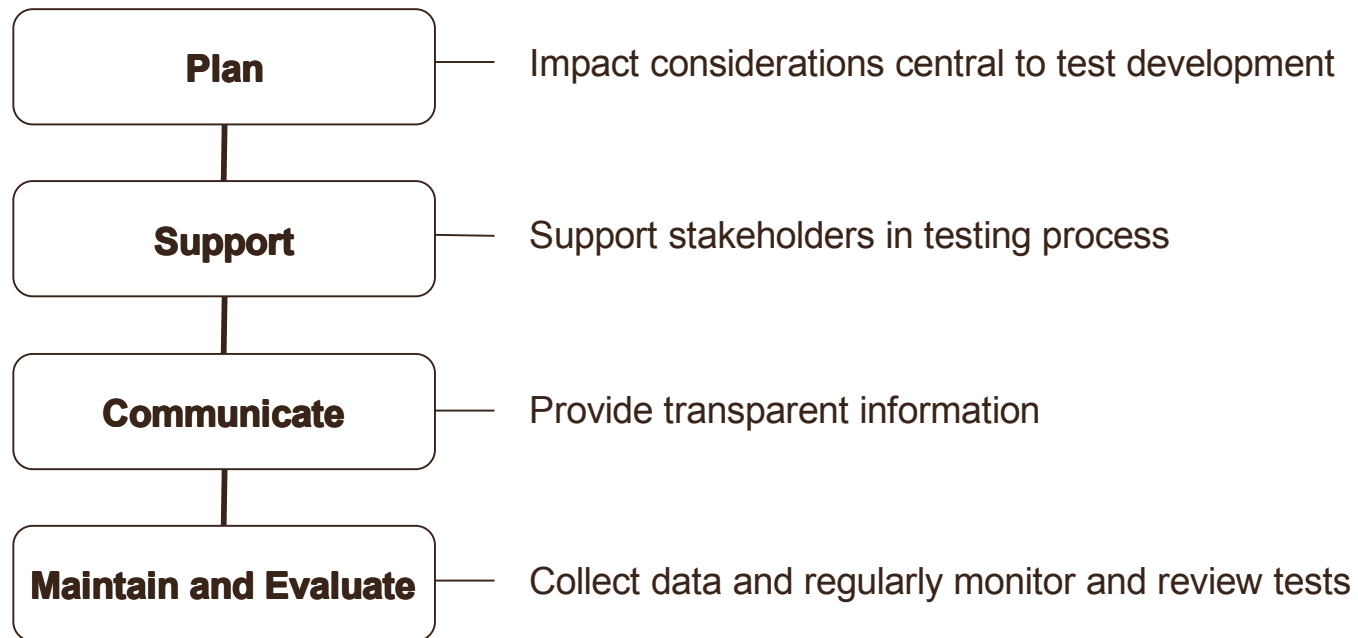## NAFLE, Beijing 2011

Impact

Impact by design

- strive to achieve positive impact on curriculum and learning

- test features that are consistent with those in instructional programmes

- an action-oriented approach to investigating impact.

Cambridge English

Impact

Impact by design

Plan	Impact considerations central to test development
Support	Support stakeholders in testing process
Communicate	Provide transparent information
Maintain and Evaluate	Collect data and regularly monitor and review tests

Very important role of Impact Studies

Impact

Ethics and fairness

- Social consequences of testing (Messich, 1989)

- Kunnan's Test Fairness Framework (2004)

- Codes of Practice (e.g. ALTE Code of Practice)

- International Language Testing Association – Code of Ethics (2000)

P | PRACTICALITY

Cambridge English

Practicality

Integral part of test usefulness and affects many different aspects of an exam

• resources necessary to produce and deliver an exam

• processing systems

• security – item-banks, exam centres, statistical malpractice checks

Practicality

Integral part of test usefulness and affects many different aspects of an exam

• test length

• availability of markers

• accessibility – exam centres

• frequency of test administration

• cost of exam

Cambridge English

NAFLE, Beijing 2011

Q QUALITY MANAGEMENT

Quality

Quality and Accountability

Quality is about achieving VRIP and fit for purpose tests consistently

- underpinning organisational and processing systems

- looking for continual improvement

- audits of processes

External international standards for Quality e.g. ISO 9001

Cambridge English

NAFLE, Beijing 2011

Quality

An action-based approach

Exam Review and Evaluation

- formal annual evaluation

- formal exam reviews at least every 5 years

Cambridge English

Quality

Underpinning organisational systems

- processing systems

- delivery systems e.g. computer-based testing

- test materials production and calibration

- item-banking systems

- examiner training and standardisation routines

Cambridge English

NAFLE, Beijing 2011

Quality

Communication and Stakeholder Support

- information and support services

- developing assessment literacy

- supporting recognising organisations

- publications and the media

- customer satisfaction surveys

Cambridge English

A framework for

Validity

Reliability

Positive **I**mpact

Practicality

Within a **Q**uality Management framework

Cambridge English

The development of a VRIPQ framework

= The Responsibilities of International Test Providers

Spolsky (1981) – *High stakes language tests should be labelled like medicine – 'use with care'*

Ensure there are ticks against

> **V**alidity ✓
>
> **R**eliability ✓
>
> **I**mpact ✓
>
> **P**racticality ✓

and a strong **Q**uality ✓ Management system

to make sure the house has strong foundations …

Cambridge English

NAFLE, Beijing 2011

Thank you –

Any questions?