

Development of Rating Scales for Performance Based Tests

Susan Jones
Lead Academic, English Language (Asia)
Trinity College London



Language testing: Judging the past or giving direction for the future

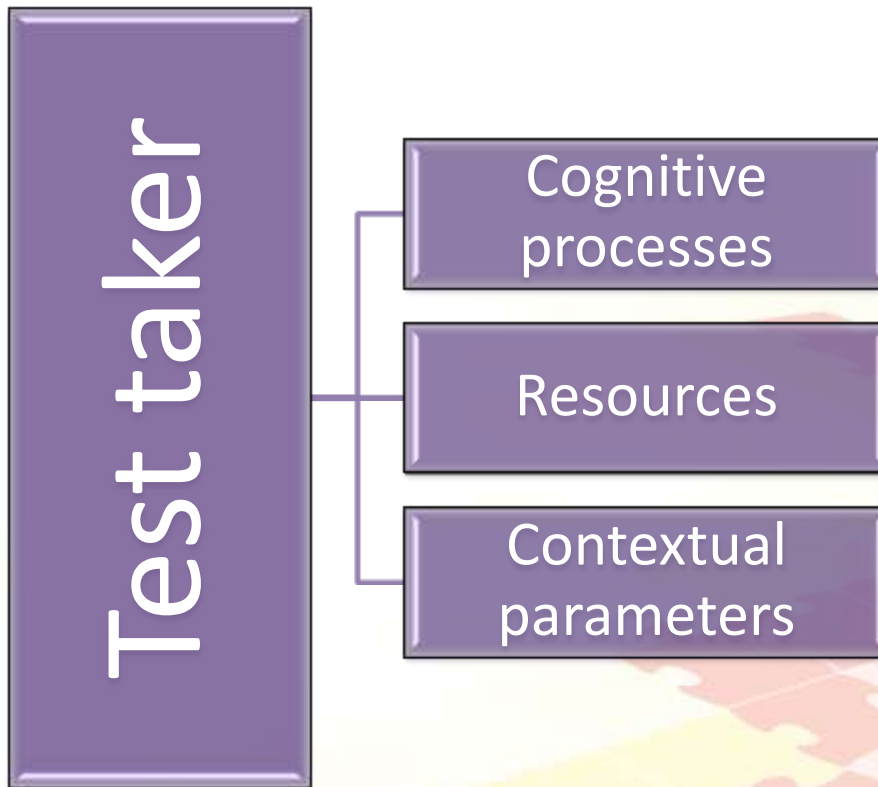


Overview

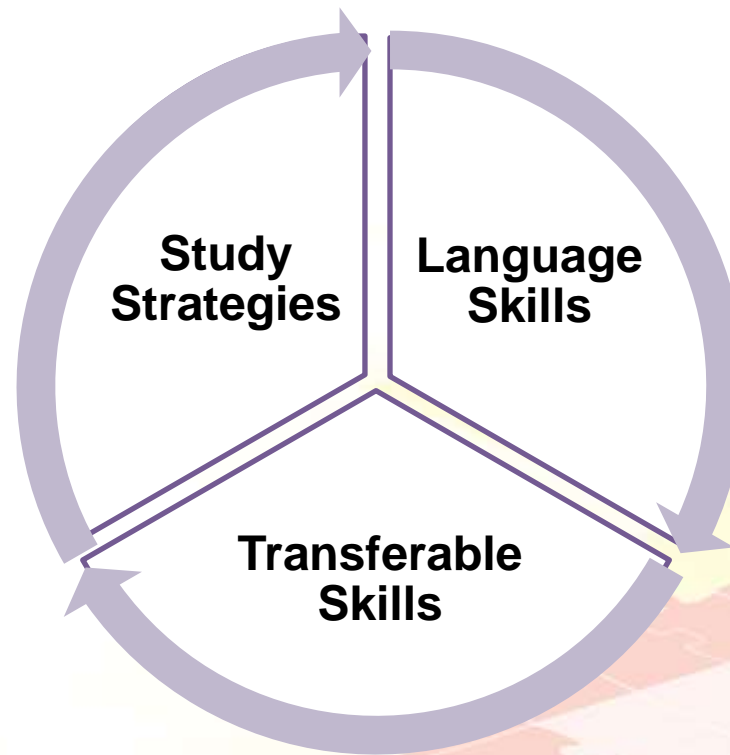
- Underlying principles of Trinity ISE exams
- Validity argument
- Cognitive validity and ISE
- How the rating scales reflect authentic, integrated communicative skills
- Differentiation in the rating scales
- Examiner training and validation

Test development – Where to start?

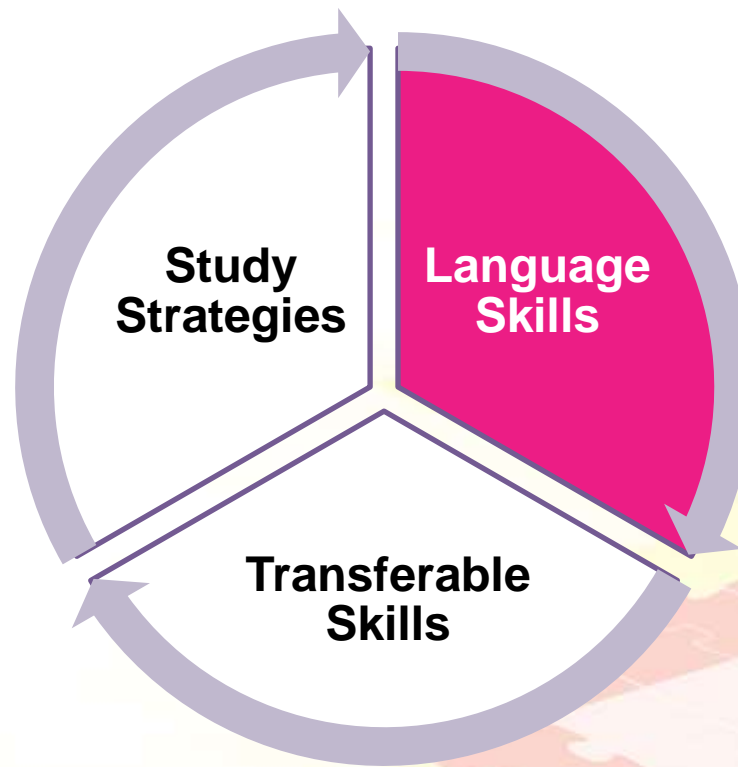
*Placing the individual test taker at the heart of the language testing agenda.
para from O'Sullivan 2011*



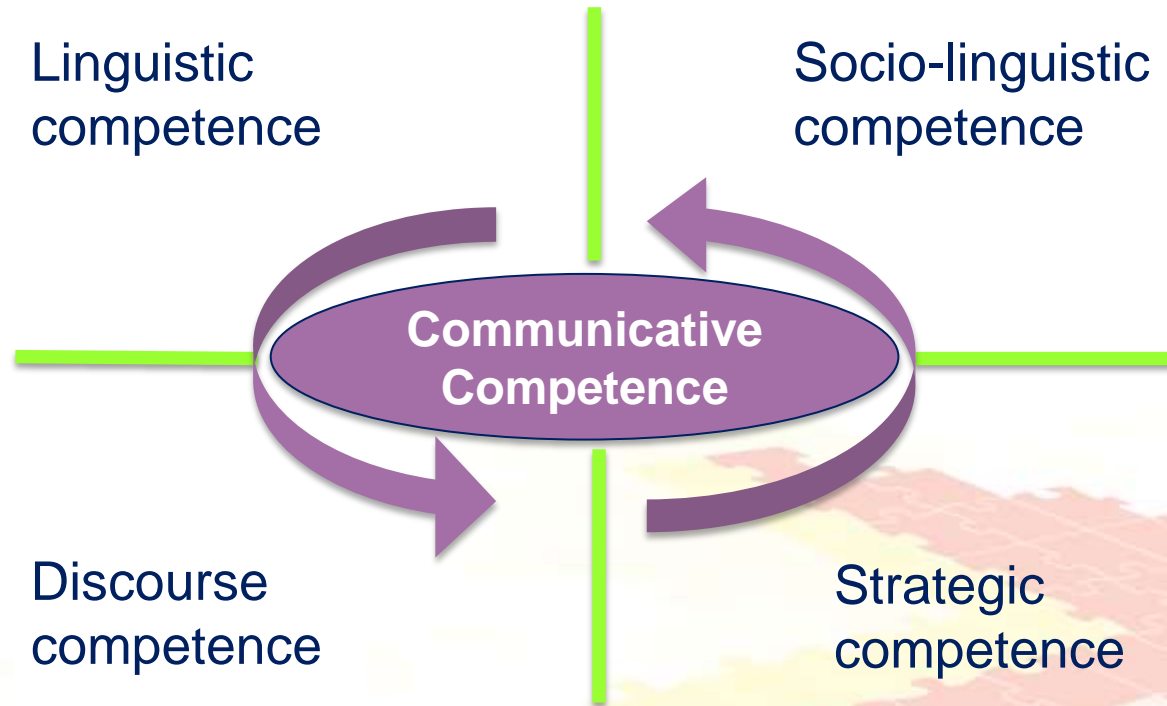
Skills students need to succeed



Skills students need to succeed

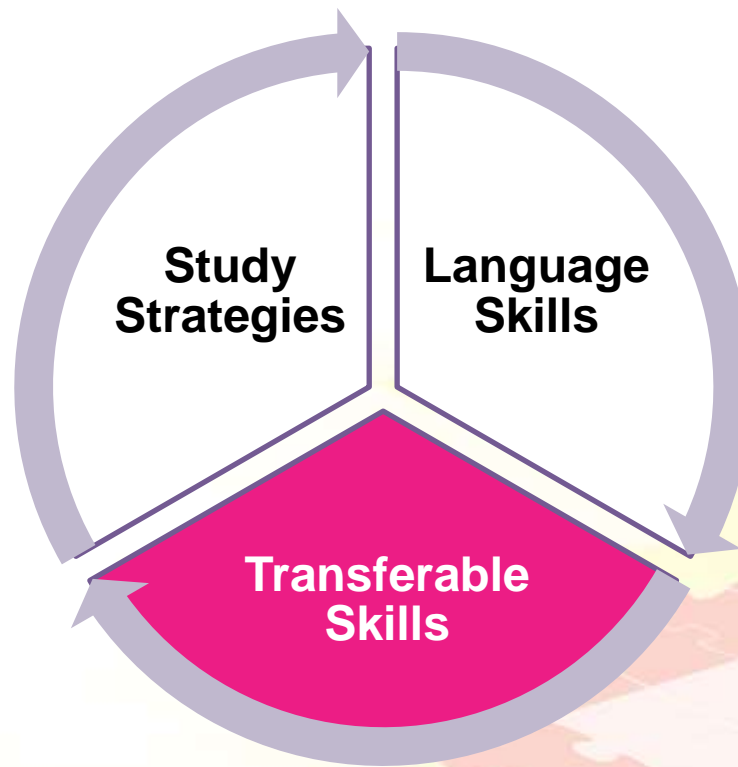


Skills students need to succeed



Bachman & Palmer (2010): communicative competence model

Skills students need to succeed



Skills students need to succeed

**Tony
Wagner**

Expert in
Residence
Innovation
Lab of
Harvard

Critical thinking & problem solving

Collaboration and leading by influence

Agility and adaptability

Initiative and entrepreneurship

Effective oral and written communication

Accessing and analyzing information

Curiosity and imagination

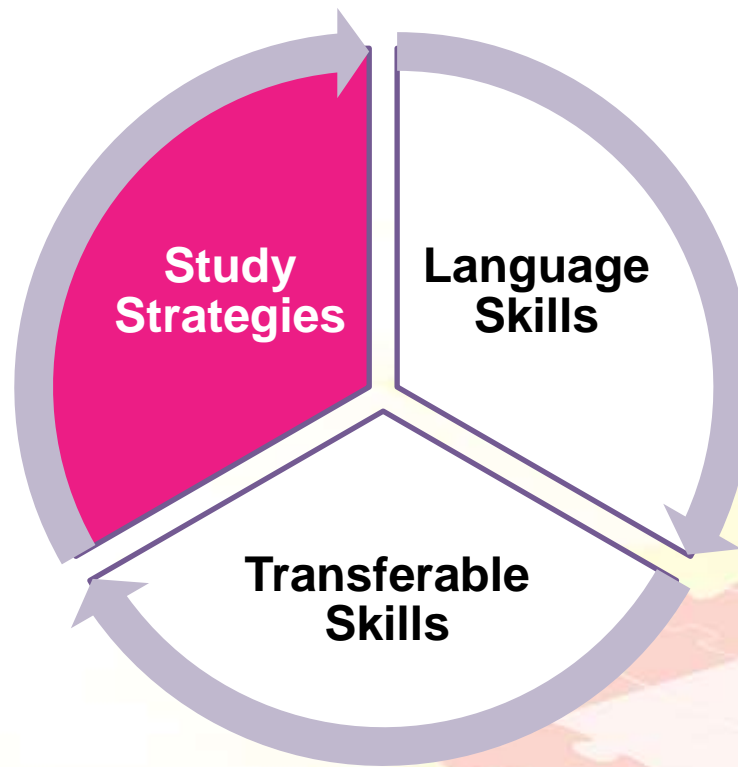
Skills students need to succeed

Global employability survey 2015:

“The most important criteria employers now use for recruitment decisions is skills profile.”



Skills students need to succeed



Skills students need to succeed

Study Strategies

Note-taking

Summarising

Synthesising

Sourcing relevant data across texts

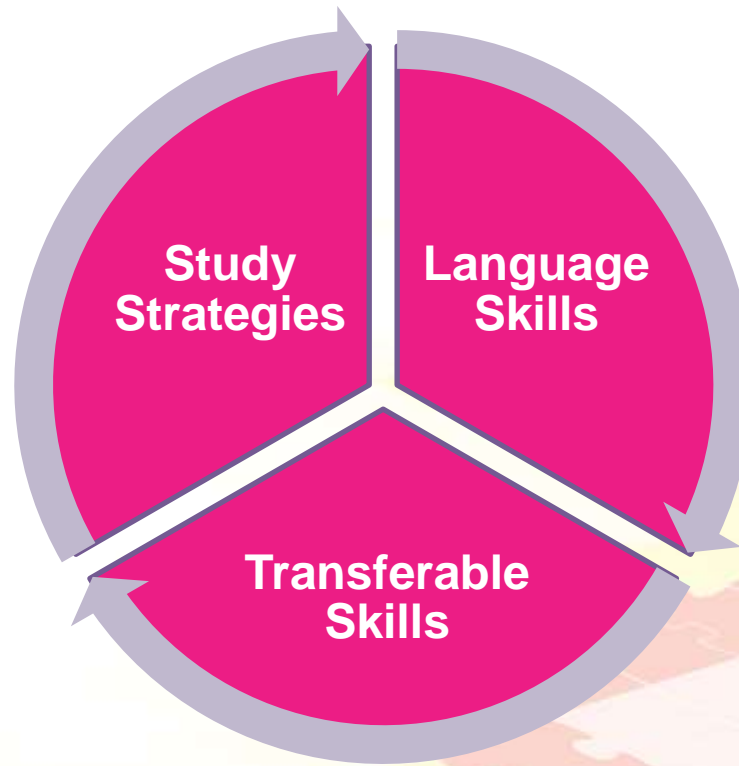
Planning

Self-checking and self-evaluation

Deducing and inferring

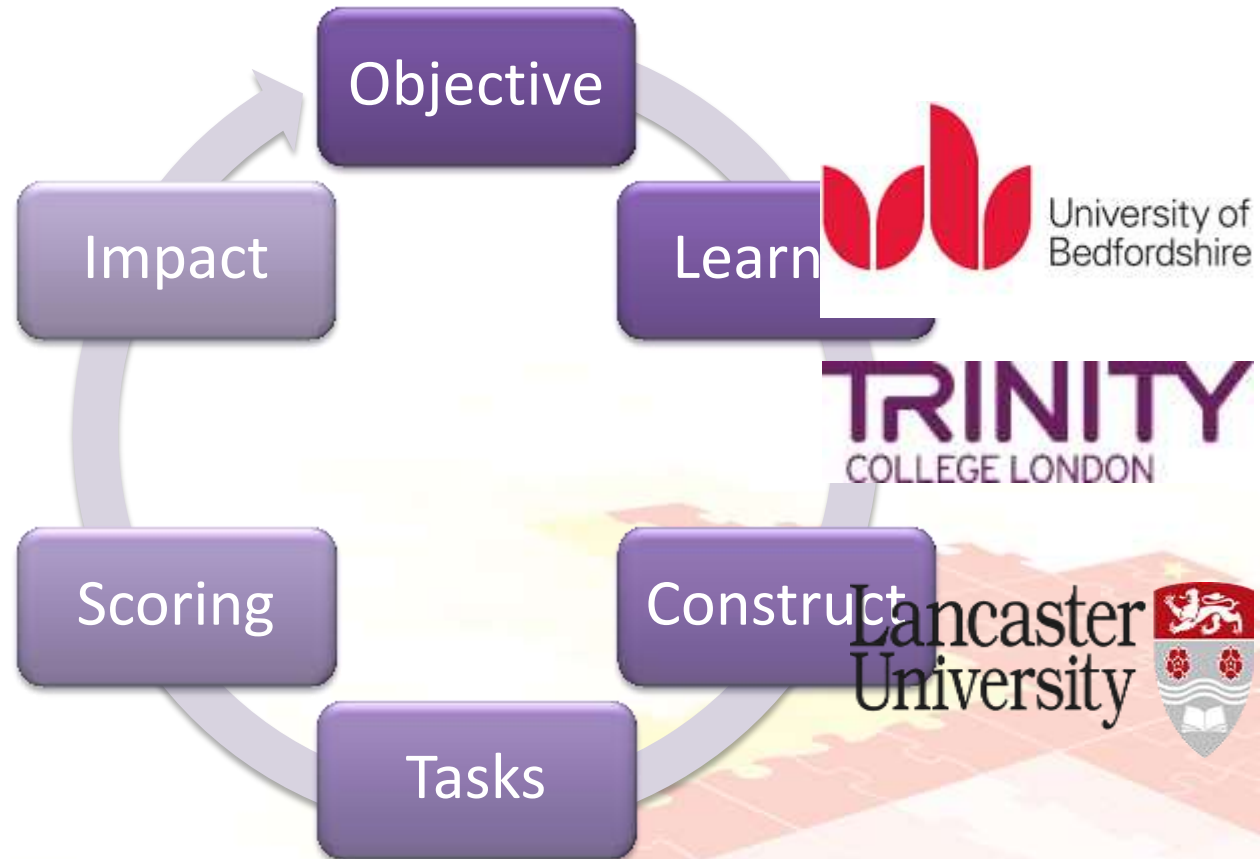
Skills students need to succeed

In an **ideal world**, a test would...



...place the test taker at the **centre** of the test.

Test design



Validity...

Table 1.3 Kinds of validity that have been proposed over the decades

Abstract	Content sampling	Divergent	Incremental	Nomological	Scoring
Administrative	Context	Domain	Indirect	Occupational	Self-defining
Aetiological	Contextual	Domain-selection	Inferential	Operational	Semantic
Artifactual	Convergent	Edumetric	Instructional	Particular	Single-group
Behavior domain	Correlational	Elaborative	Internal	Performance	Site
Cash	Criteria	Elemental	Internal test	Postdictive	Situational
Circumstantial	Criterion	Empirical	Interpretative	Practical	Specific
Cluster domain	Criterion-oriented	Empirical-judgemental	Interpretive	Predictive	Statistical
Cognitive	Criterion-related	Essential	Intervention	Predictive criterion	Status
Common sense	Criterion-relevant	Etiological	Intrinsic	Predictor	Structural
Communication	Cross-age	External	Intrinsic content	Prima Facie	Substantive
Concept	Cross-cultural	External test	Intrinsic correlational	Procedural	Summative
Conceptual	Cross-sectional	Extratest	Intrinsic rational	Prospective	Symptom
Concrete	Cultural	Face	Item	Psychological & logical	Synthetic
Concurrent	Curricular	Factorial	Job analytic	Psychometric	System
Concurrent Criterion	Decision	Faith	Job component	Quantitative face	Systemic
Concurrent Criterion-related	Definitional	Fiat	Judgemental	Rational	Theoretical
Concurrent true	Derived	Forecast true	Known-groups	Raw	Theory-based
Congruent	Descriptive	Formative	Linguistic	Relational	Trait
Consensual	Design	Functional	Local	Relevant	Translation
Consequential	Diagnostic	General	Logical	Representational	Translational
Construct	Differential	Generalized	Longitudinal	Response	Treatment
Constructor	Direct	Generic	Lower-order	Retrospective	True
Construct-related	Discriminant	Higher-order	Manifest	Sampling	User
Content	Discriminative	In situ	Natural	Scientific	Washback
Content-related					

Validity – not a simple concept
Newton and Shaw (2014)



Validity arguments

Construct

- The test assesses what it claims to assess

Face

- Test takers feel that the test does what it says it does

Scoring

- The measurement of the performance is valid

Criterion related

- Predictive validity and concurrent validity

Context

- Performance conditions are appropriate

Content

- Appropriateness of linguistic and content demands of text

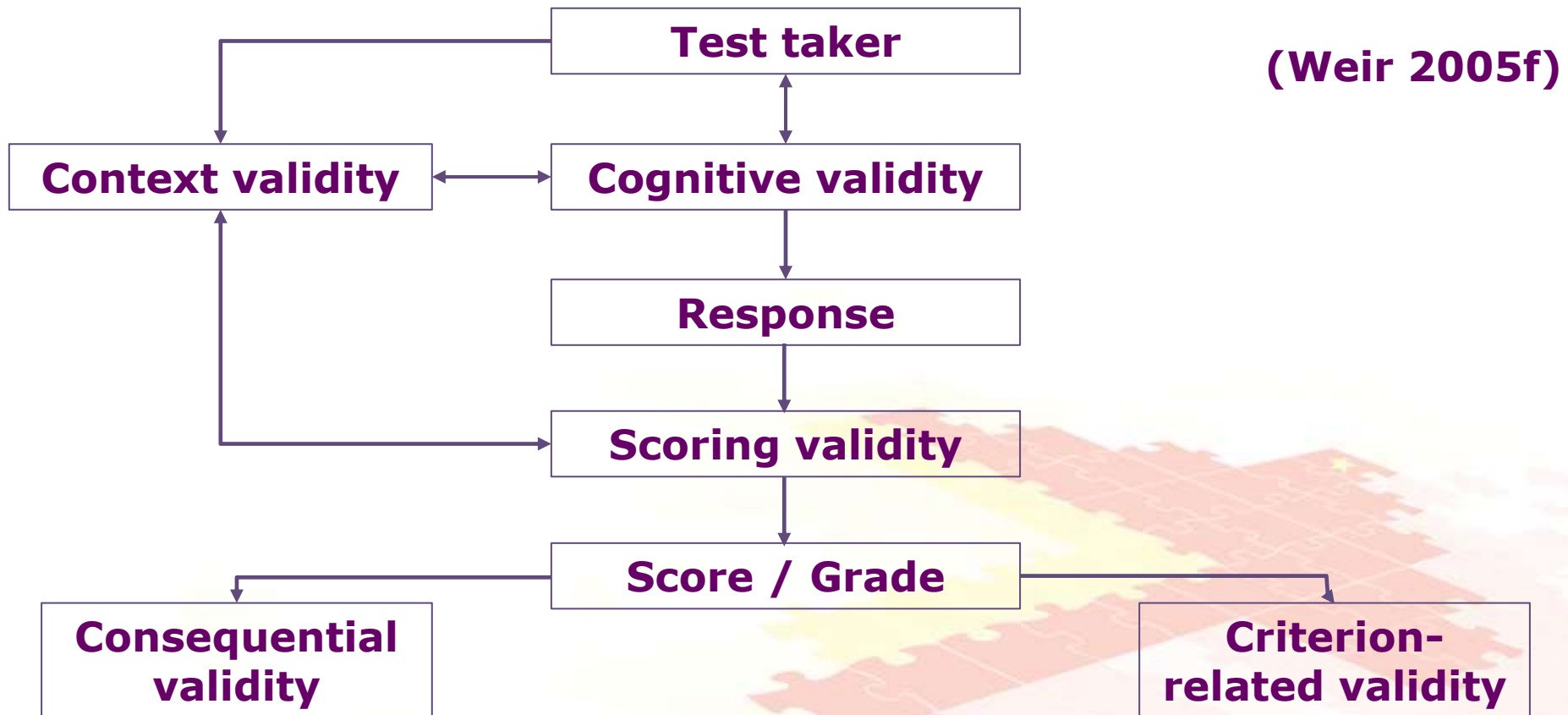
Cognitive

- Engages performance of the target cognitive processes

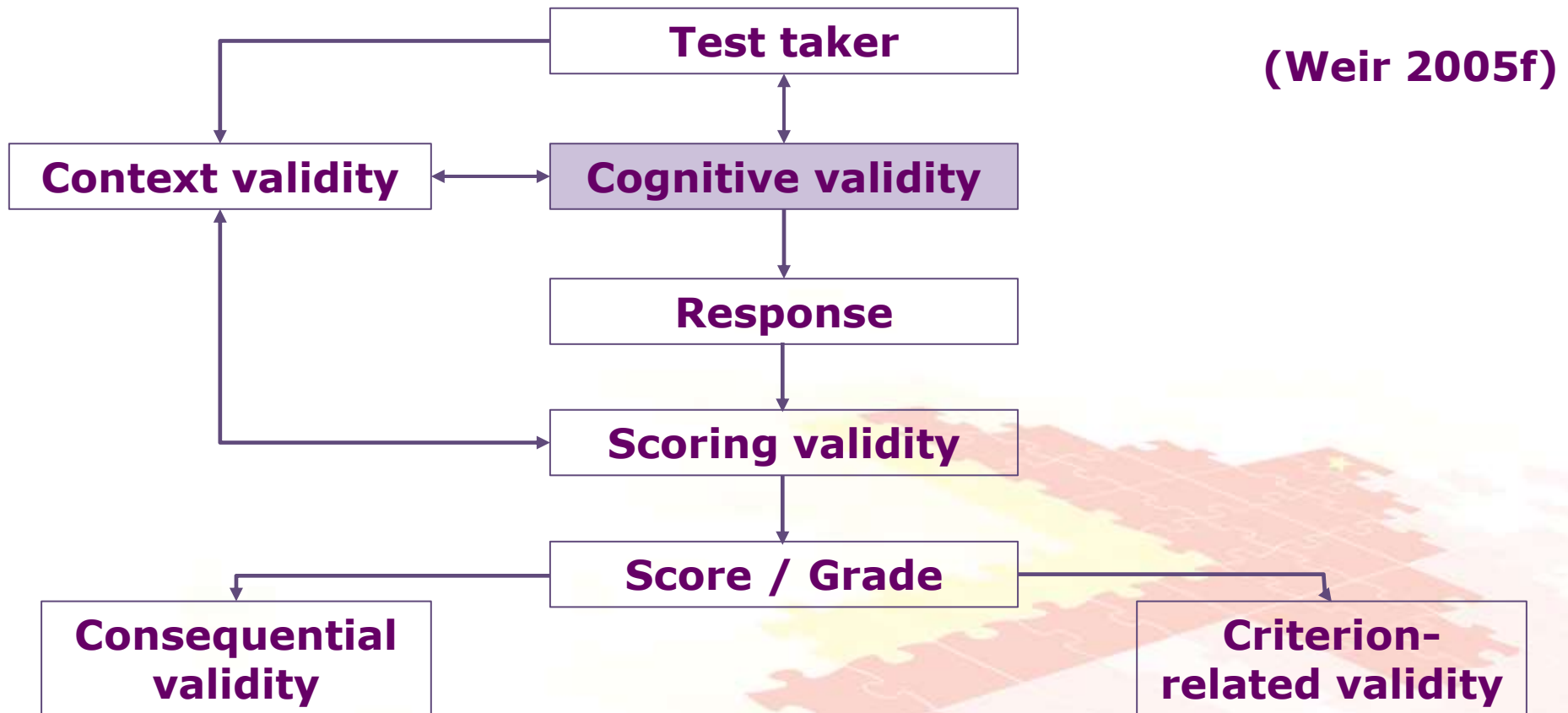
Consequential

- Effect of exam following exam (can be long term)

Cognitive validity – Socio-cognitive framework



Cognitive validity – Socio-cognitive framework



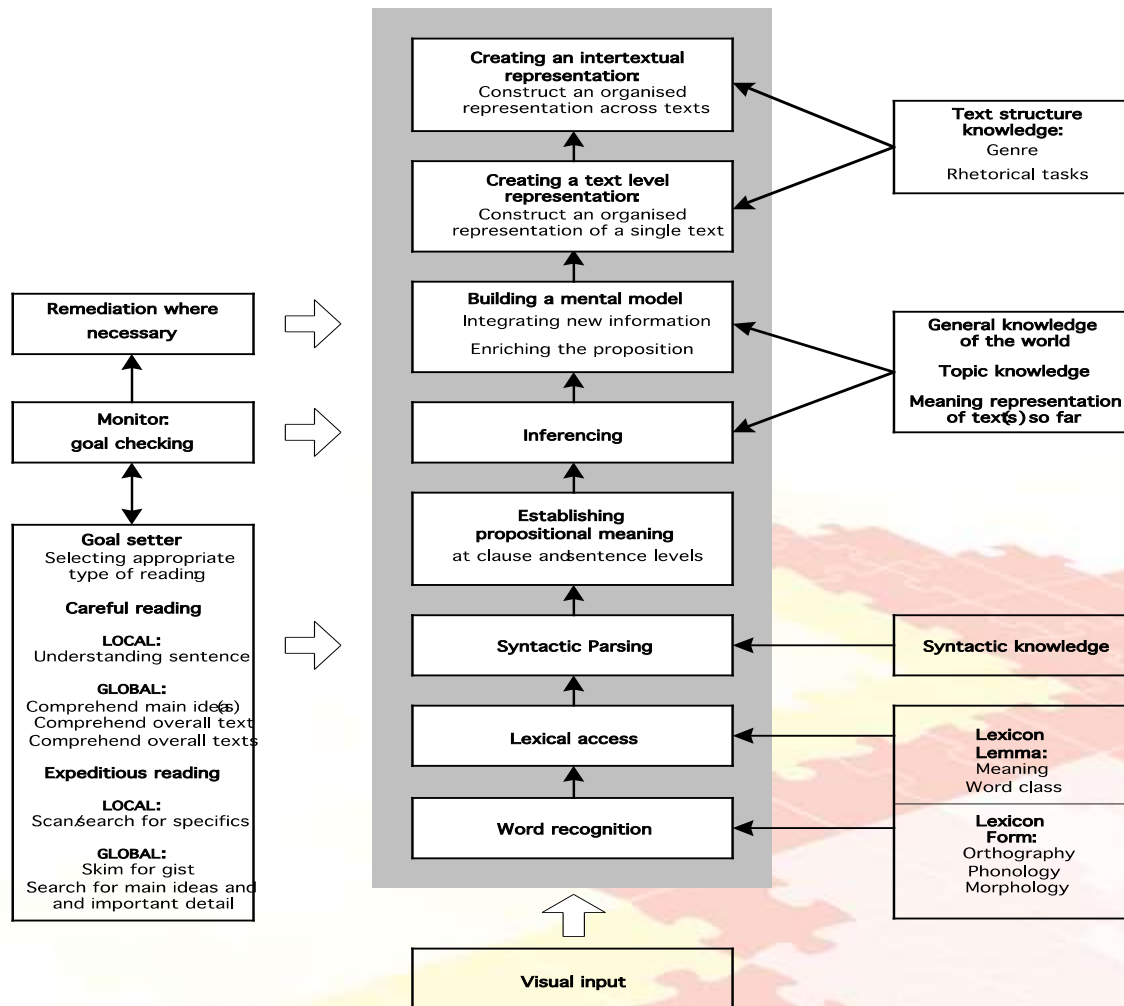
Trinity ISE

Integrated Skills in English

	ISE Foundation A2	ISE I B1	ISE II B2	ISE III C1
Reading & Writing	4 tasks with the same structure across all four levels: 2x Reading tasks, 1x Reading into writing task, 1x Writing task <i>2 hours</i>			
Speaking & Listening	2x Speaking tasks 2x Listening tasks 13 minutes	2x Speaking tasks 2x Listening task 18 minutes	3x Speaking tasks 1x Listening task 20 minutes	3x Speaking tasks 1x Listening task 25 minutes

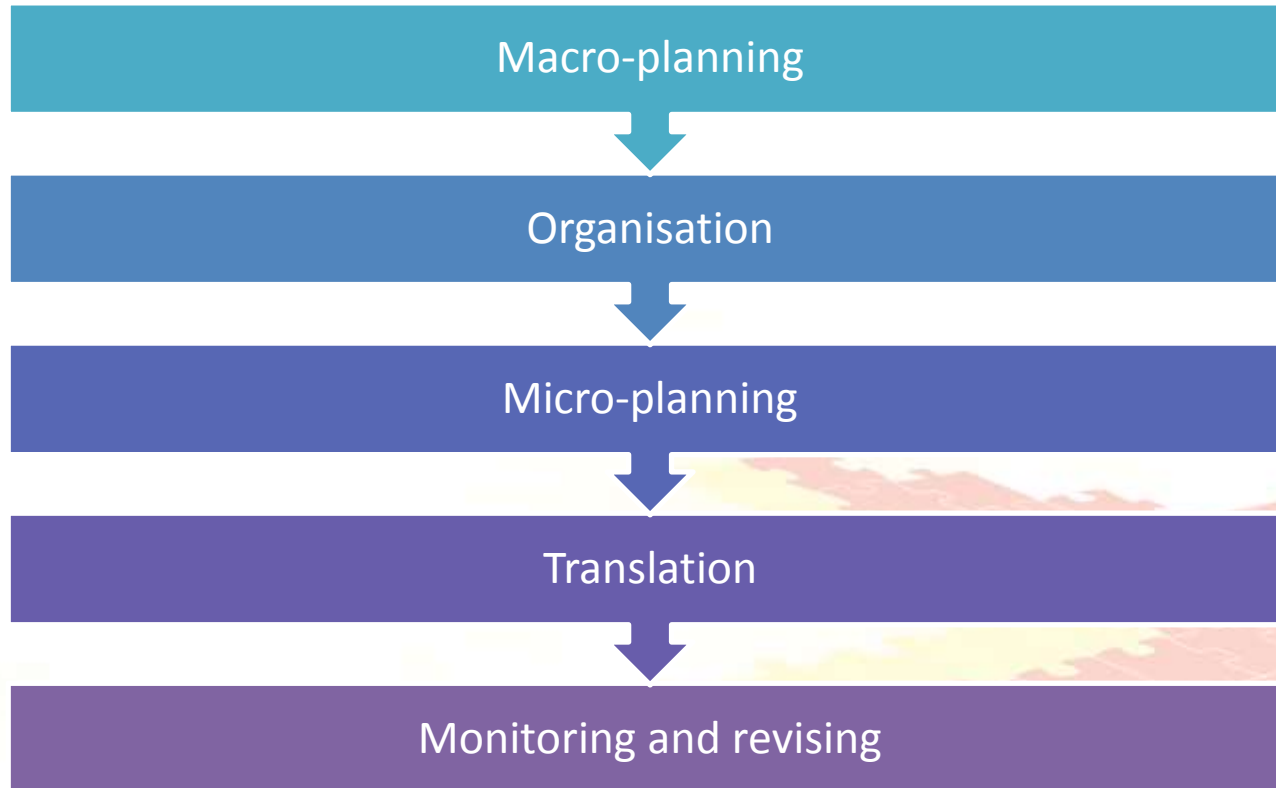
Reading - Cognitive processing

Behind the scenes (Khalifa and Weir, 2009)



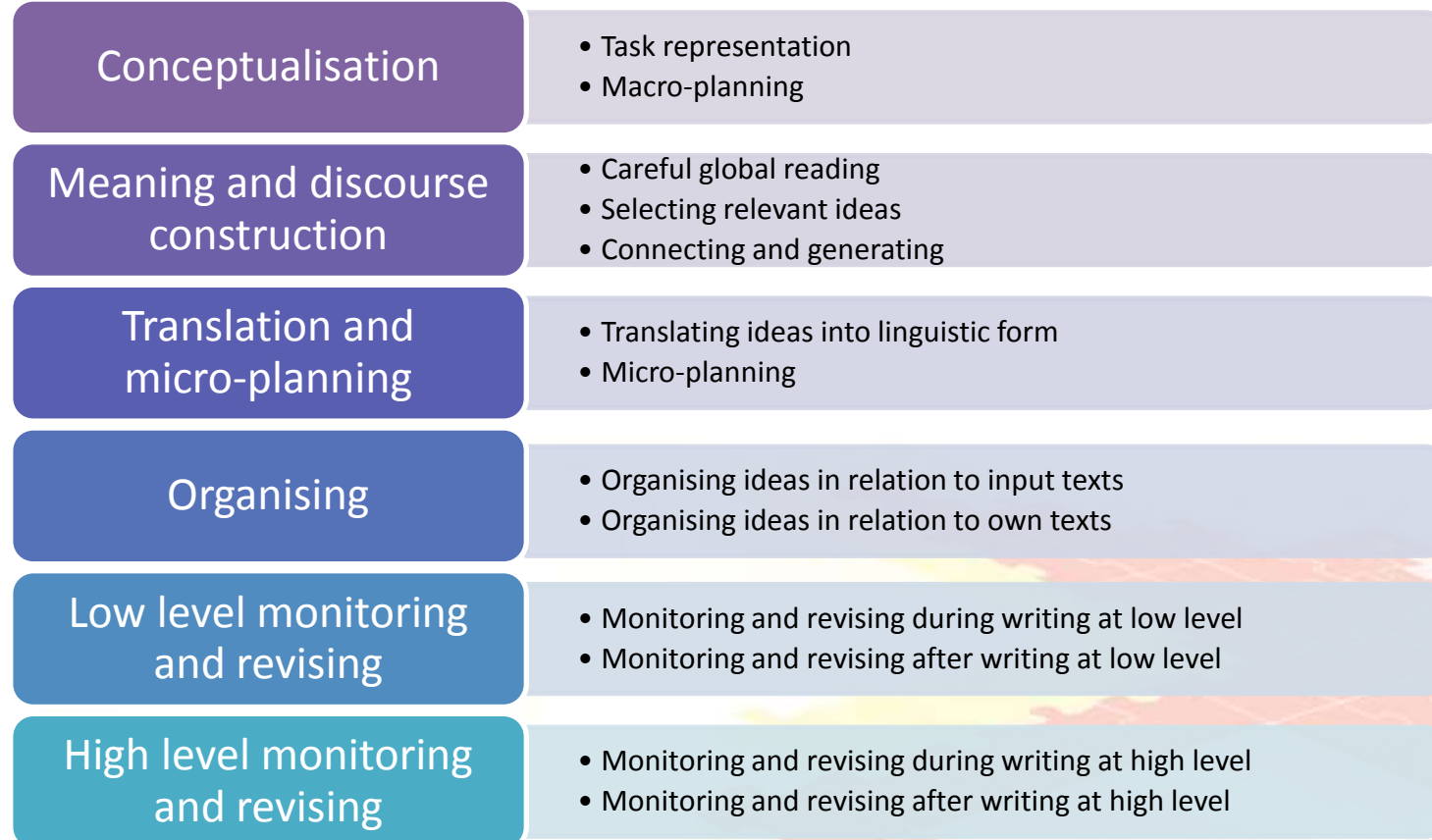
Writing - Cognitive processing

Behind the scenes (Shaw and Weir, 2007)



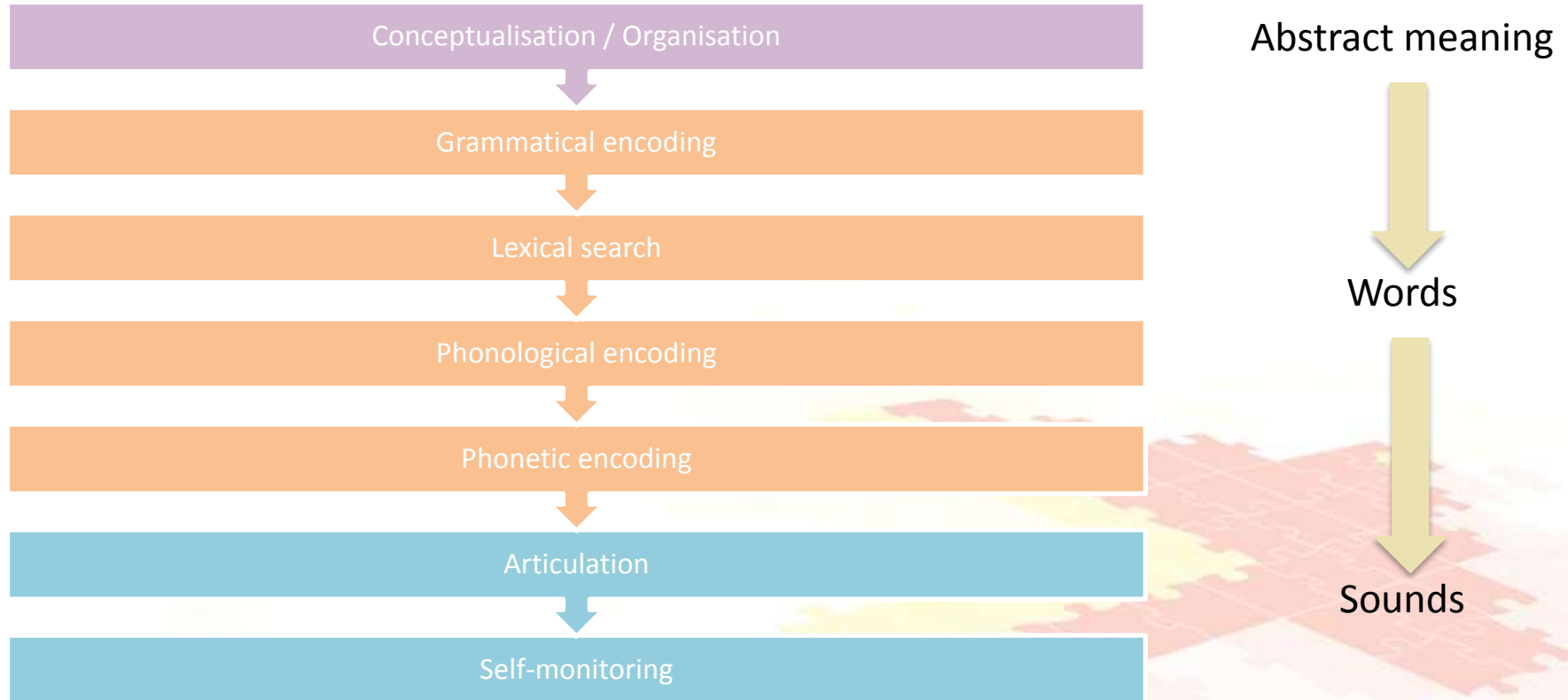
Reading into writing - Cognitive processing

Behind the scenes (Chan, 2013)



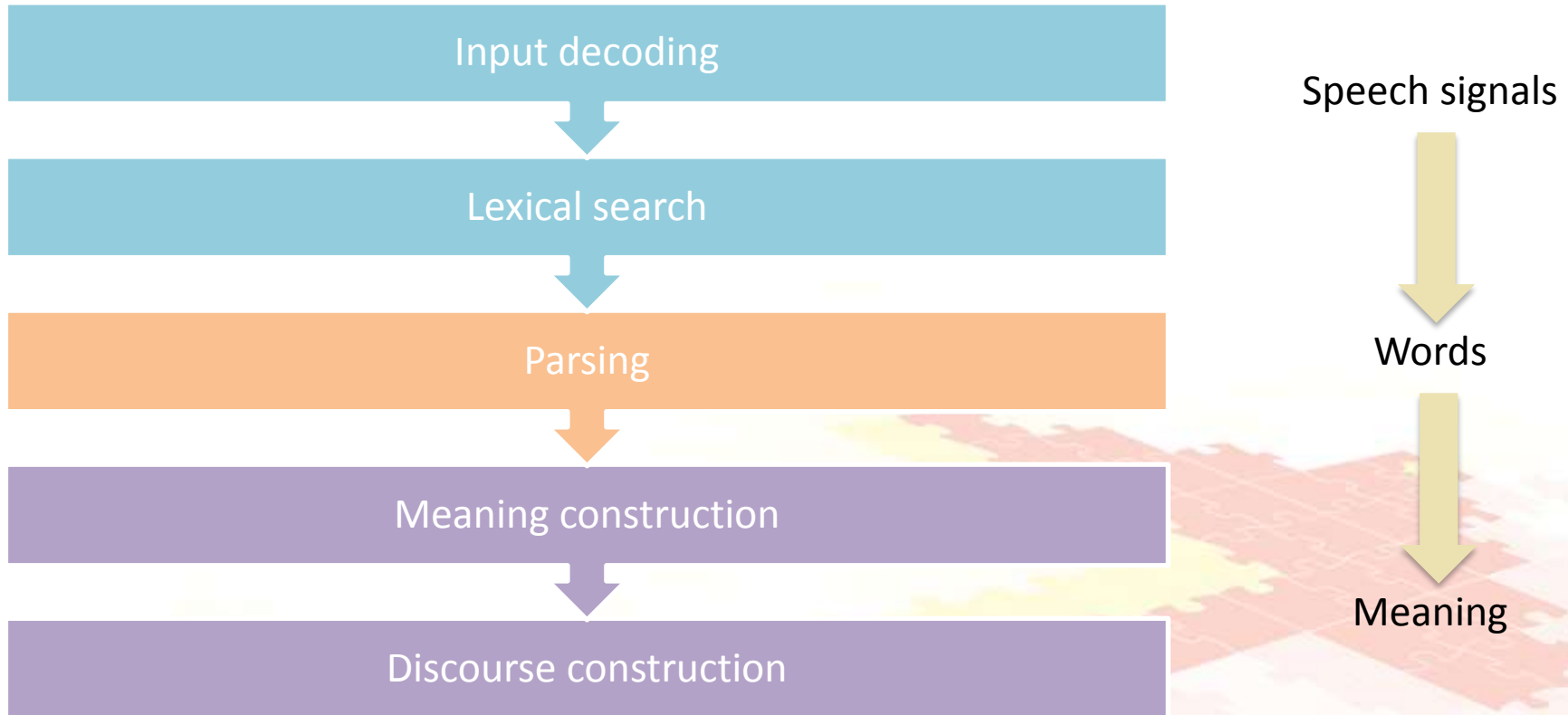
Speaking - Cognitive processing

Behind the scenes (Field, 2011)




Listening - Cognitive processing

Behind the scenes (Field, 2008)

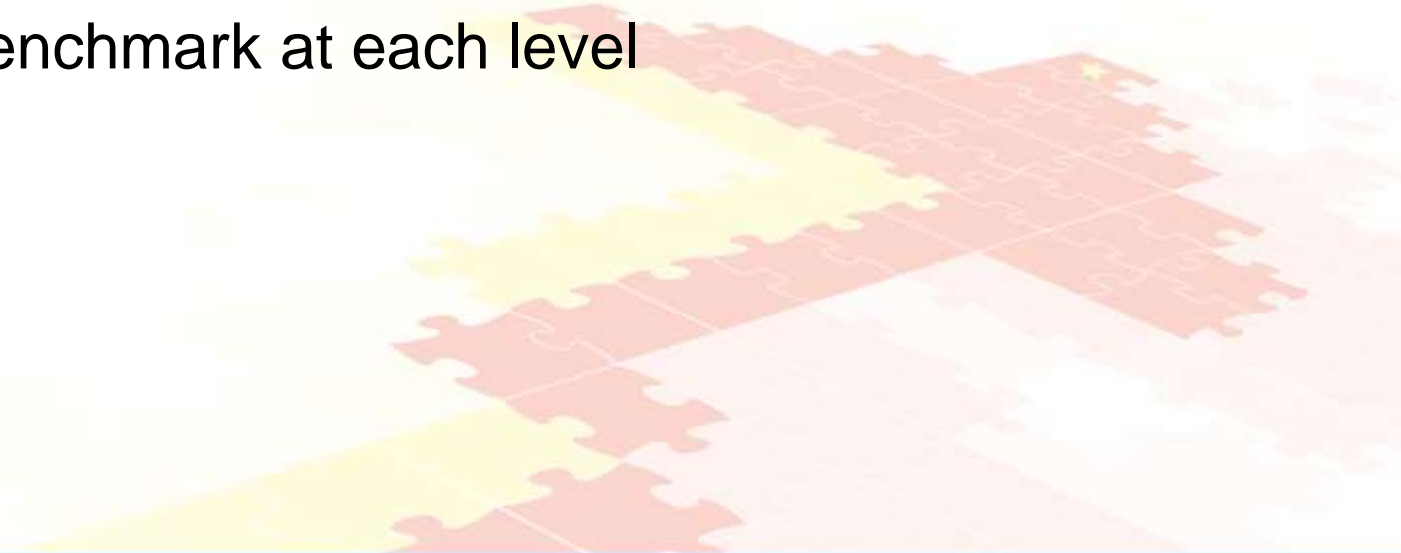


Rating scales – R/W

- Reading
 - Main idea / Gist
 - Sentence (main and supporting arguments, fact and opinion)
 - Word or phrase level
 - Reading into writing
 - **Reading for writing**
 - Task fulfilment
 - Organisation and structure
 - Language Control
 - Extended writing
 - Task fulfilment
 - Organisation and structure
 - Language Control
- 

Rating scales – S/L

- Speaking
 - Communicative effectiveness
 - **Interactive listening**
 - Language control
 - Delivery
- Independent listening
 - CEFR benchmark at each level

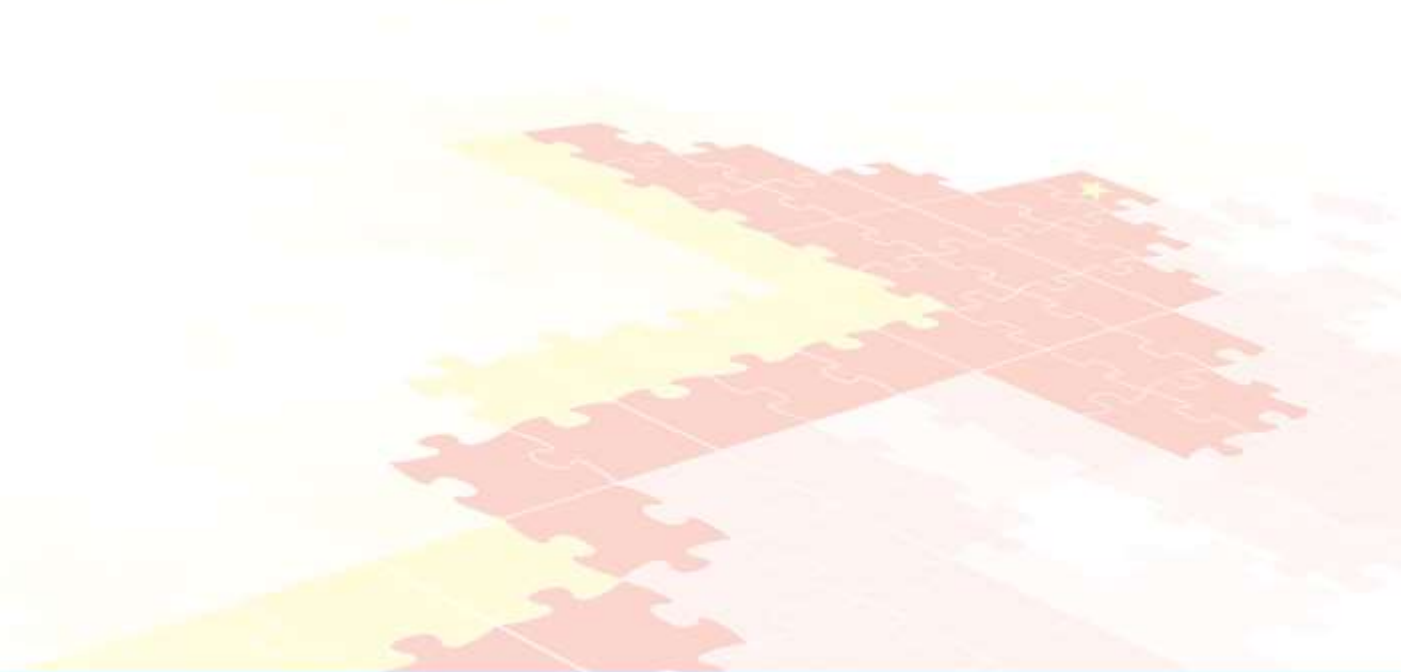


Differentiation in rating scales

- A2 – Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
- B1 – Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics, which are familiar, or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
- B2 – Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
- C1 – Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

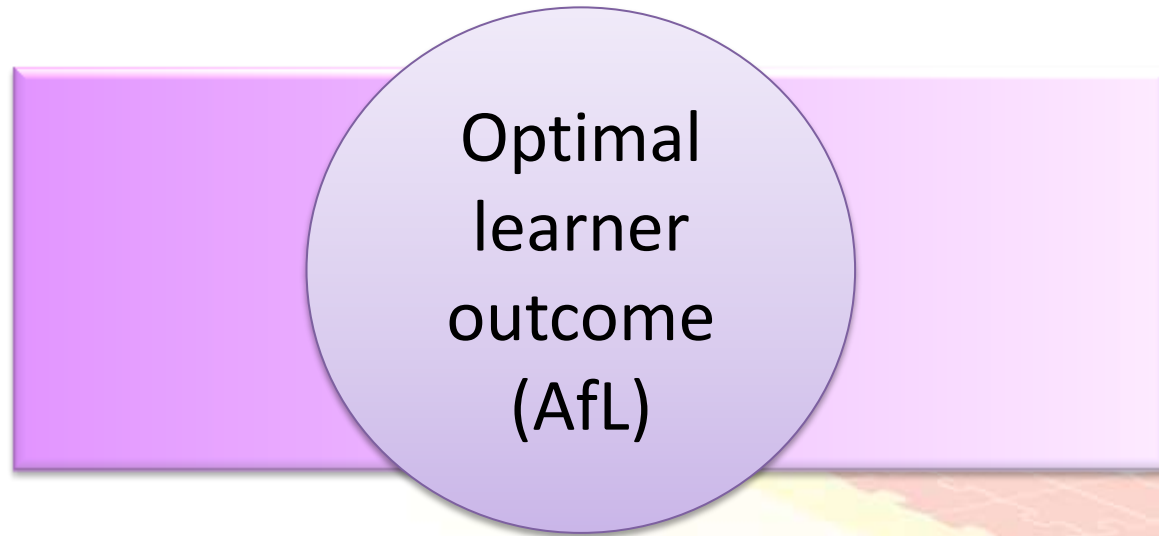
Examiner training and validation

- Validity vs validation
- Examiners
 - Training and standardisation
 - Monitoring
- Robust awarding processes



Assessment for learning

TEACH



TEST

Thank you!

Any questions, please contact
susan.jones@trinitycollege.com

