

## Argumentum ad scientiam: Language testing for legal students and professionals

Ivana Vidaković  
BAAL TEA SIG, 11<sup>th</sup> June 2013



### Overview and aims

- Background to ESP test validation
- The International Legal English Certificate (ILEC)
- The framework used for its validation
- ILEC test validation: assumptions, the evidence gathered, methodology, findings and validity arguments



## ESP test validation: Criticisms

- EAP/ESP testing criticised for a simplistic view of validity (Fulcher 1999):
  - Focus on: authenticity and specificity of task content and format
  - Sampling and representativeness
  - Not enough focus on the construct/ability and (the validity and reliability of) the inferences drawn from test scores
- Test validation or task analysis frameworks re-thought for an EAP/ESP test context:
  - Messick's (1989) by Fulcher (1999)
  - Bachman & Palmer's (1996) by Douglas (2000)
  - Useful, but not sufficiently comprehensive and/or readily usable
 So...



## ESP test validation 'how-to': from a **claim and framework to scientia & argument**

- Select a framework for test validation which:
  - Is comprehensive - covering all aspects of test validity
  - Is practical
  - Allows validation along criteria relevant for language tests in general and ESP tests in particular (authenticity, specificity, the role of background knowledge)
- Make an inference/assumption about your test
  - Interpretive argument (Kane 1990)
- Gather quantitative and/or qualitative evidence from test/item analysis, assessment and content specialists (e.g. panels, focus groups, questionnaires)
- Make a validity argument
  - To justify or modify your inferences/assumptions (Kane 1990)



## ILEC: Ensuring validity of an ESP test

The test: The International Legal English Certificate (ILEC)

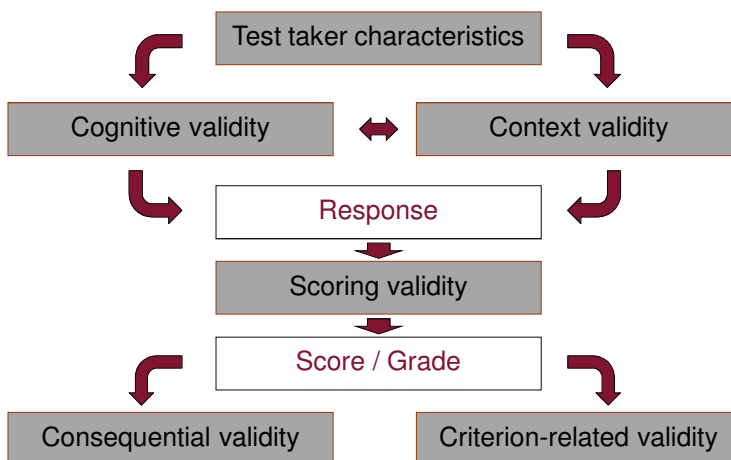
- A test for law students and lawyers
- Assesses English language skills in the context of international commercial law
- Set at B2 and C1 levels of the CEFR
- Recognised by a large number of colleges, universities, as well as law firms and government bodies

The framework for test validation: socio-cognitive (Weir 2005)

- Language ability = language knowledge + context-dependent language use
- Language use is affected by social and cognitive factors
- Comprehensive, practical; a general framework, but could be applied to ESP tests



## A socio-cognitive framework for test validation (Weir 2005)







## Six key questions

- 1) How are the characteristics of test takers catered for by this exam?
- 2) Are the cognitive processes required to complete the task appropriate?
- 3) Are the characteristics of the test tasks and the test administration fair to the test takers?
  - Is the context acceptable to the candidates and content specialists (in terms of familiarity and relevance)?
  - Are real-life features and conditions incorporated in the tasks?
- 4) How far can we depend on the scores of the test?
- 5) What external evidence is there outside of the test scores themselves that the test is doing a good job?
- 6) What effects does the test have on its stakeholders?



## The needs of ILEC test takers

1. Needs analysis and creating a sample exam:
  - Assessment specialists (Cambridge English)
  - Content specialists – TransLegal (qualified lawyers, law lecturers, specialists in civil law, providers of legal English services), TransLegal lawyers across Europe – consulted on real-life tasks, texts, task purpose...
2. Judging the authenticity of the sample exam and the demand for the exam:
  - Global legal community – content specialists:
    - Decision makers at law firms, Universities and international lawyer associations
    - Qualified lawyers, teachers of legal English, law lecturers
3. Test production process:
  - Assessment specialists – the whole process, including CIS (candidate information sheet) data analysis/monitoring
  - TransLegal as specialist *content vetters* towards the end of the process


 

## Context validity: ILEC Reading and Listening papers



<b>Context validity-Task:</b>	<b>Context validity-Text:</b>
Purpose (✓)	Discourse mode ✓
Rubric (✓)	Text length (✓)
Task types (✓)	Writer-reader relationship (✓)
Known criteria ✓	Nature of information & topic (✓)
Weighting (✓)	Content knowledge demands (✓)
Time constraints (✓)	Linguistic demands (✓)

(✓) partially done

- ILEC needs analyses (Cambridge Assessment & TransLegal)
- *ILEC sample test: Market research study* (Chamberlain & Robinson 2004)
- *ILEC trial test study* (Thigh 2005)




© UCLES 2013

## Contextual features are appropriate for law students and legal professionals

	<b>Needs analysis</b>	<b>Sample test study</b>	<b>Trial test study</b>	<b>Test production</b>
<b>Type of data</b>	Quan, qual	Quan, qual	Quan, qual	Item writer guidelines
<b>Method</b>	Focus groups, interviews, q'aires	Q'aires	Q'aires, Shadowing	Pretesting and editing
<b>Research participants</b>	Content specialists	Content specialists	Test takers: students & lawyers	Vetting by content specialists



© UCLES 2013



## Authenticity: Trial 2005


Questionnaire Statements about ILEC	Legal professionals (60)	Legal students (226)
Reading paper topics are authentic	58 (92%) strongly agree & agree 3 (4.8%) missing data	191 (85%) strongly agree & agree 22 (10%) missing data
Texts and language in the Reading paper are authentic	56 (89%) strongly agree & agree 3 (4.8%) missing data	187 (83%) strongly agree & agree 21 (9%) missing data
Listening tasks are authentic	52 (83%) strongly agree & agree 5 (8%) missing data	167 (74%) strongly agree & agree 24 (11%) missing data




## Timing: Trial 2005

**Enough time to complete each part of each component? Yes (majority)**

- Feedback from candidates and instructors (questionnaires)
- Shadowing a set of candidates during the test
- Quantitative analysis of candidate item responses and uncompleted or omitted items.

 CAMBRIDGE ENGLISH  
Language Assessment  
Part of the University of Cambridge




## Cognitive validity: ILEC Reading paper


**Cognitive validity:**


- (✓) Cognitive processes
- (✓) Resources :
  - Content knowledge
  - Language knowledge  
(e.g. discursal, lexical)

(✓) partially done

*ILEC Reading: Exploring cognitive validity evidence*  
(Vidakovic & Khalifa 2010)


© UCLES 2013 

 CAMBRIDGE ENGLISH  
Language Assessment  
Part of the University of Cambridge



## ILEC reading like workplace/academic reading: cognitive processes, reading types

- **Research subjects:** 10 law students (B2 & C1)
- **Methodology:** **a)** Immediate introspection: a checklist with cognitive processes and reading types (QUAN), **b)** Focus group interviews after the test – to dig deeper & wider (QUAL)
- **Findings:**
  - The assumption confirmed: a variety of cognitive processes (word to text level) and reading types (e.g. fast, careful) elicited. They were previously attested in work/academic contexts.
  - *Task focus* and *gapping* influence processes and reading types
  - Task focus on *text content* and an *engaging rhetorical structure* encourage a more authentic approach to reading: legal reasoning & reading most of the text, creating mental maps of textual organisation
  - Drawing on *knowledge of linguistic features* typical of legal texts helped arriving at answers

© UCLES 2013 



## Scoring validity: ILEC Reading and Listening papers

### Scoring validity:

- ✓ Item-level statistics (difficulty, facility, discrimination): pretests and live tests
- ✓ Reliability – internal consistency coefficient Alpha (Based on regular Classical and IRT analyses)
- ✓ Differential item functioning (DIF)
- ✓ Marker reliability: RITCME, a list of appropriate responses
- ✓ Grading: cut-off scores based on pre-determined ability levels and difficulty of live test items

*Differential item functioning in ILEC Reading and Listening (Vidakovic 2011)*



## No candidates are (dis)advantaged

**Data:** Responses of 898 candidates to May 2009 ILEC R and L papers

### Methodology:

- DIF in terms of gender, L1 (Romance vs. non-Romance) and work experience (WE)/no work experience as a legal professional (NWE)
- BILOG-MG software, Marginal Maximum Likelihood Ratio
- Content analysis of test items (only in terms of word origin)

### Findings:

- The assumption above confirmed for gender and L1 groups
- Lexical similarities *do not consistently advantage* Romance candidates, as word use & word environment awareness also matter
- DIF for WE: 6 R and 2 L items favoured WE candidates, 3 R items favoured NWE candidates. *DIF overlap:* 2/3 items favour NWE candidates due to their Romance L1 (44% are Rom. L1 speakers)
- Tasks requiring localised reading (word-sentence level) engender more DIF (66% items) than tasks with global reading (across paragraph(s), text) - 22% DIF items – context helps?





## Criterion validity: ILEC Reading and Listening papers

### Criterion- related validity:

- X Cross test comparability
- ✓ Equivalence with different versions of the same test (item banking, the Cambridge ESOL Common Scale: all pretested items linked to the Common Scale through anchor items)
- X Comparability with external standards



## Context validity: ILEC Writing test

### Context validity-Task:



- Purpose (✓) \*
- Rubric (✓)
- Task types (✓)
- Known criteria (✓)
- Weighting (✓)
- Time constraints (✓)

### Context validity-Text:

- Discourse mode (✓)
- Text length (✓)
- Writer-reader relationship (✓)
- Nature of information & topic (✓)
- Content knowledge demands (✓)
- Linguistic demands :  
Lexical ✓ \*\*

\*(✓) partially done, \*\*✓ done


- Needs analysis, *Sample test: Market research study* (Chamberlain & Robinson 2004), *ILEC trial test study* (Thigh 2005)
- *Lexical resources in legal and general English writing across L2 proficiency levels: a corpus-informed study* (Vidakovic & Barker 2010)



 CAMBRIDGE ENGLISH  
Language Assessment  
Part of the University of Cambridge 

## Contextual features are appropriate for law students and legal professionals

Questionnaire Statements about ILEC	Legal professionals (60)	Legal students (226)
Writing paper topics are authentic	41 (65%) strongly agree & agree 21 (33%) missing data	63 (27.9%) strongly agree & agree 159 (70.4%) missing d.
Writing tasks are authentic	38 (60%) strongly agree & agree 21 (33%) missing data	60 (27%) strongly agree & agree 159 (70.4%) missing d.


Thigh (2005)

© UCLES 2013 

 CAMBRIDGE ENGLISH  
Language Assessment  
Part of the University of Cambridge 

## Legal character of written responses reflected in lexis and lexical bundles

- **Data:** extended written responses to Writing test papers of 160 ILEC (B2/C1), 80 FCE (*Cambridge English: First*; B2) and 80 CAE (*Cambridge English: Advanced*; C1) candidates
- **Methodology:** WordSmith, Vocabulary Profiler for single-word frequency analyses; QUAL analysis of the function of lexical bundles (multi-word units). Comparison within and across the exams.
- **Findings:**
  - ILEC responses more lexically sophisticated than FCE & CAE (7.5% academic words at each level vs. 2% in FCE and 4% in CAE)
  - Sentences much longer in ILEC responses – greater structural complex.
  - More legal words in ILEC responses (5% vs. < 1% in FCE/CAE data)
  - Lexical bundles are more frequent and varied in ILEC responses
  - Discourse structuring bundles – very prominent in ILEC responses
  - Domain of language possibly influences acquisition of words & multi-word units at B2 & C1 levels, reducing differences between the levels.

© UCLES 2013 



## Scoring validity: ILEC Speaking test

### Scoring validity:

- ✓ Assessment criteria/scales
- ✓ Rating procedures and conditions:
  - Training & standardisation of Oral Examiners (OE-s): RITCME
  - The Interlocutor framework
  - Pair format: analytic and holistic rating
- ✓ Statistical analyses (of OE performance):
  - MFRA\* analysis for marking trials and standardisation videos
  - Correlations of analytic to holistic marks for examiner monitoring

- Vidakovic, I. and E. Galaczi (2009) ILEC Speaking: revising assessment criteria and scales. *Cambridge ESOL: Research Notes* Issue 35: 29-34.
- Annual ILEC Speaking reports



## The revised ILEC Speaking assessment scales function adequately

- The aim of the revision: to align the scales closely with an external criterion - the CEFR
- *A-priori* validation based on QUAN and QUAL evidence
- Methodology: 34 raters, 12 candidates, MFRA, q'aires
- Findings:
  - Assessment criteria: within acceptable range of variability in difficulty and fit (scores neither predictable nor too predictable)
  - Oral Examiners – within acceptable limits of severity and fit (consistency in awarding marks)
  - Some raters had difficulty assessing the appropriateness of use of legal vocabulary
  - Grammar & vocabulary hardest to assess: getting the balance right between a rich (general and legal) vocabulary and a much less accurate and varied grammar



## ILEC test validation: A summary

Overview of ILEC test validation with a focus on evidence, methodology, findings and validity arguments:

- Usually a combination of quantitative and qualitative data
- The evidence confirmed the starting assumptions about the test's validity and some of it fed into rater training
- The research studies and the use of the socio-cognitive framework shows the need for further research:
  - a) to further explore the investigated aspects of validity
  - b) to explore the aspects of validity that have not undergone a-posteriori validation (e.g. the context validity of the ILEC Speaking paper, the cognitive validity of the ILEC Listening, Speaking and Writing papers, the consequential validity of ILEC, a-posteriori validation of Speaking and Writing assessment scales)



## Future ESP research

- Investigate cognitive processes to determine how test takers with and without specific background knowledge perform on and approach ESP tasks (and to complement considerable focus in ESP literature on the influence of background knowledge on test scores)
- How to make specificity and authenticity measurable?
  - Refining Clapham's (1996) questionnaire based on the TMC (Test Method Characteristics) Scale or O'Sullivan's (2006) instrument based on Weir's Socio-Cognitive Framework
  - Comparing test data with corpora of real-life documents (creating e-corpora with spoken and written discourse)
- Training assessment specialist raters to rate discipline-specific vocabulary or training content specialists to do so



Thank you for listening.

Any questions?

Vidakovic.I@cambridgeenglish.org



### Authenticity: Sample test study 2004

	ILEC contains the kind of authentic test material appropriate in a test of legal English	
	Agree	Disagree
Legal organisations & Uni departments (51)	75%	6%
Legal English instructors (15)	80%	7%



## Authenticity: Trial 2005

Questionnaire Statements about ILEC	Legal professionals (60)	Legal students (226)
Reading paper topics are authentic	58 (92%) 3 (4.8%) missing data	191 (85%) 22 (10%) missing data
Texts and language in the Reading paper are authentic	56 (89%) 3 (4.8%) missing data	187 (83%) 21 (9%) missing data
Writing paper topics are authentic	41 (65%) 21 (33%) missing data	63 (27.9%) 159 (70.4%) missing d.
Writing tasks are authentic	38 (60%) 21 (33%) missing data	60 (27%) 159 (70.4%) missing d.
Listening tasks are authentic	52 (83%) 5 (8%) missing data	167 (74%) 24 (11%) missing data

© UCLES 2013



## Why have ESP tests?

- Linguistic features vary by discipline  
(see Tiersma (1999) for legal English)
- Language performances vary with context and task  
(Tarone 1985, Ellis 1989, O'Sullivan 2006)
- Language ability = language knowledge + context-dependent language use (e.g. Weir 2005)
- Face validity (Fulcher 1999, Davies 2001) – effect on the construct:
  - To motivate test takers to do their best
  - To enable a deeper and more authentic engagement with tasks (Alderson 2000, Weir 2005)
- Impact on test users
- To better sample the relevant ability and to better predict field-specific linguistic performance

© UCLES 2013