



Asset Languages – evaluation of pilot Manual

Neil Jones, Tamsin Walker, Karen Ashton



Context of study

- Our purposes in doing the work reported here:
 - To develop practical procedures for constructing a complex multilingual assessment framework.
 - To contribute more generally to the development of a sound framework for the validation of Cambridge ESOL language assessments.
- Our purpose in offering a case study based on this work:
 - To contribute to making the manual a more effective and practical tool.
- NB our use of the manual has been critical, selective and pragmatic.

Asset Languages in outline

- Produced under contract to DCSF
- Complex framework:
- 25 languages (modern foreign languages, community languages)
- Three contexts of learning (Primary, Secondary, Adult),
- Reading, writing, listening and speaking separately tested.
- Two assessment strands: certificated external assessment, and more informally accredited teacher assessment

www.assetlanguages.org.uk



Asset Languages and the national languages strategy

- National Languages Strategy launched in 2002 to address serious underperformance in languages.
- The Languages Ladder:
 - “a new voluntary recognition system to complement existing national qualifications frameworks and the Common European Framework” (DfES 2002).



Remit of Asset Languages

- Provide a framework for lifelong learning: i.e. a *learning ladder* of accessible targets.
- Relate to meaningful proficiency levels, i.e. comparability between languages in terms of proficiency, not learning effort.



The Languages Ladder

- A document produced and owned by DCSF
- Can-do descriptions of 6 stages subdivided into 3 grades per stage
- Interpreted by Asset Languages as broadly equivalent to CEFR
- Articulated in terms of *learning stages*, not *levels achieved*



Languages Ladder – example summary stage description

Listening, Preliminary – grades 4–6

On *completing* this stage, you should be able to understand standard speech relating to a range of predictable everyday matters, providing that it is spoken clearly and directly. You should be able to recognise the difference between past, present and future events and be familiar with simple forms of the verb tenses.



CEFR levels & Languages Ladder stages

Languages Ladder:
Learning stages

CEFR:
Proficiency levels

	<u>Intermediate 9</u>	<u>B1 Threshold</u>
<i>Working at Intermediate</i>	<u>Intermediate 8</u>	<i>Working towards B1</i>
	<u>Intermediate 7</u>	
	<u>Prelim 6</u>	
	<u>Prelim 6</u>	<u>A2 Waystage</u>
<i>Working at Prelim</i>	<u>Prelim 5</u>	<i>Working towards A2</i>
	<u>Prelim 4</u>	
	<u>Breakthrough 3</u>	
	<u>Breakthrough 3</u>	<u>A1 Breakthrough</u>
<i>Working at Breakthrough</i>	<u>Breakthrough 2</u>	<i>Working towards A1</i>
	<u>Breakthrough 1</u>	

Development and validation of objective tests (Reading, Listening)

	Design, Specification, item writing	Scale Construction	Standard Setting	Research, review
Approach	ESOL tests as model  (cf CEFtrain) Reference to CEFR, LL scales for guide to appropriate tasks and linguistic/functional level	IRT: Pretesting including vertical link, common-person linking The template scale 	NC/CEFR/LL teacher ratings collected at pretesting	KA's PhD: German, Urdu, Japanese reading Reading grid, Sources of Difficulty studies Cross-language comparison workshops (qualitative, quantitative); Plurilingual can-do data
Use of Manual	Post-hoc selective application of Chapter 4 (Specification)		External validation (Chap 6) NO operational task-centred standard setting (Chap 5)	

Development and validation of performance tests (Speaking, Writing)

	Design, Specification, item writing	Standard setting, rater training, standardisation	Research, review
Approach	ESOL tests as model Reference to CEFR, LL scales. Focus on task as a major determinant of level	English exemplars as a model for setting standards across languages Other language exemplars benchmarked to CEFR	Cross-language verification studies using multilingual informants Cross-language comparison workshop (quantitative); Plurilingual can-do data
Use of Manual	Post-hoc selective application of Chapter 4 (Specification)	CEFR familiarisation exercises, reference to CoE performance exemplars (French, German) CEFR scales to add substance to LL descriptors	



Use of the Manual: Familiarisation

- Useful – we have used the CEFR to give meaning to the Languages Ladder.
- Direct application to standardising raters, also useful orientation for item writers.
- However, item writers need to be made aware that the apparent CEFR level of a text may *not* equal the difficulty of items written for it.



Use of the Manual: Specification

- Comments based on post-hoc study for purposes of case study.
- How much of the content description required is relevant to defining levels?

Specification: Form A9, Listening (Preliminary / A2)

	Reference	Description (Prelim paper)	Description useful in defining level?	Ref. useful?
Situations, content categories, domains	Table 5, CEF 4.1 p.48	Personal + Public; Shopping, meal out, watching TV, travel, phone message,		No Only a sample, relates to much more practical use of language, categories too fine.
Communication themes	Lists, CEF 4.2 p.52	Shopping, food and drink, travel, weather, hobbies and interests, TV	Duplication with previous – should content categories be removed from previous?	Yes
Communicative tasks	Lists, CEF 4.3 p.53-6	Understand simple facts: times; common expressions; numbers; jobs; likes/dislikes	Y – but learner doesn't need to understand likes/dislikes, even though in text	Not particularly Only a sample p.54 "indefinitely large in number" 'real world' not applicable to tests



Specification

- Forms A8 to A18 (CEFR Chapter 4 - language in use) don't relate so clearly to level.
- Complexes of contextual parameters don't necessarily identify a level because they are *manipulated* to create an appropriate general level of linguistic challenge. Treating them separately doesn't capture this.
- Forms A19 onwards deal with CEFR Chapter 5 (Language competence) and are much easier to work with.
- Better to put these first – gives a more logical progression from classroom learning to extrapolating to real world tasks.



Standardisation

- Performance skills well treated.
- The treatment of standard setting is idiosyncratic:
 - task centred approach = “standard setting”, an essential first step
 - learner-centred approach = “external validation”, an optional verification process
- Cross-language comparison is not treated but is crucial



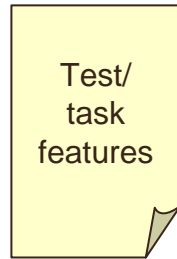
General comments on manual

- Who is the audience? - Language testers with varying resource and expertise.
- The Manual: “not a guide to how to construct valid tests” - but this isn’t true:
- Its central concern – how to make valid criterion-referenced interpretations of test performance (to the CEFR) – is the central problem of test validity.
- Thus aligning tests to the CEFR can’t be a one-off exercise - should be integrated into every stage of the design and administration cycle.
- The Manual has educative potential for quality assurance, particularly for smaller or less well-resourced operations.
- cf p125: composing the internal report “could contribute to the on-going process of continuously improving the quality of the examination involved, e.g. construction of the test, test development, feedback for item writers and raters, etc.”
- Thus the organization of the Manual should relate more explicitly to the scheme of such a testing cycle

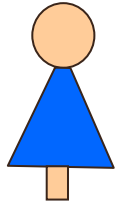
Test construction

Theory-based validity

Context validity



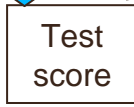
Learner features



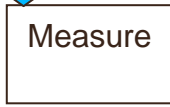
1



2



3



What to observe?
How?

How can we score what we observe?

Scoring validity

Are scores consistent and interpretable?

Measurement validity

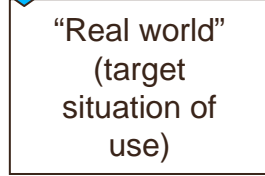
Evaluation

Generalization

Extrapolation

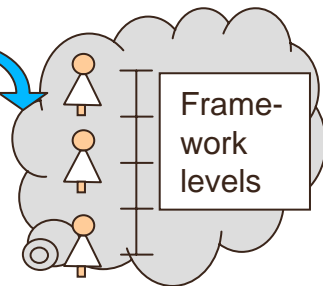
Alignment

4

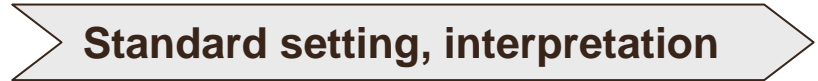
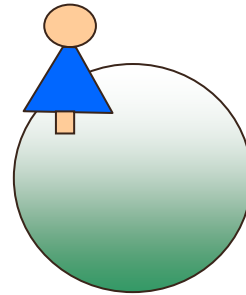


Does the test score reflect the candidate's actual ability?

5



How does the specific learning/testing context relate to a more general proficiency framework?



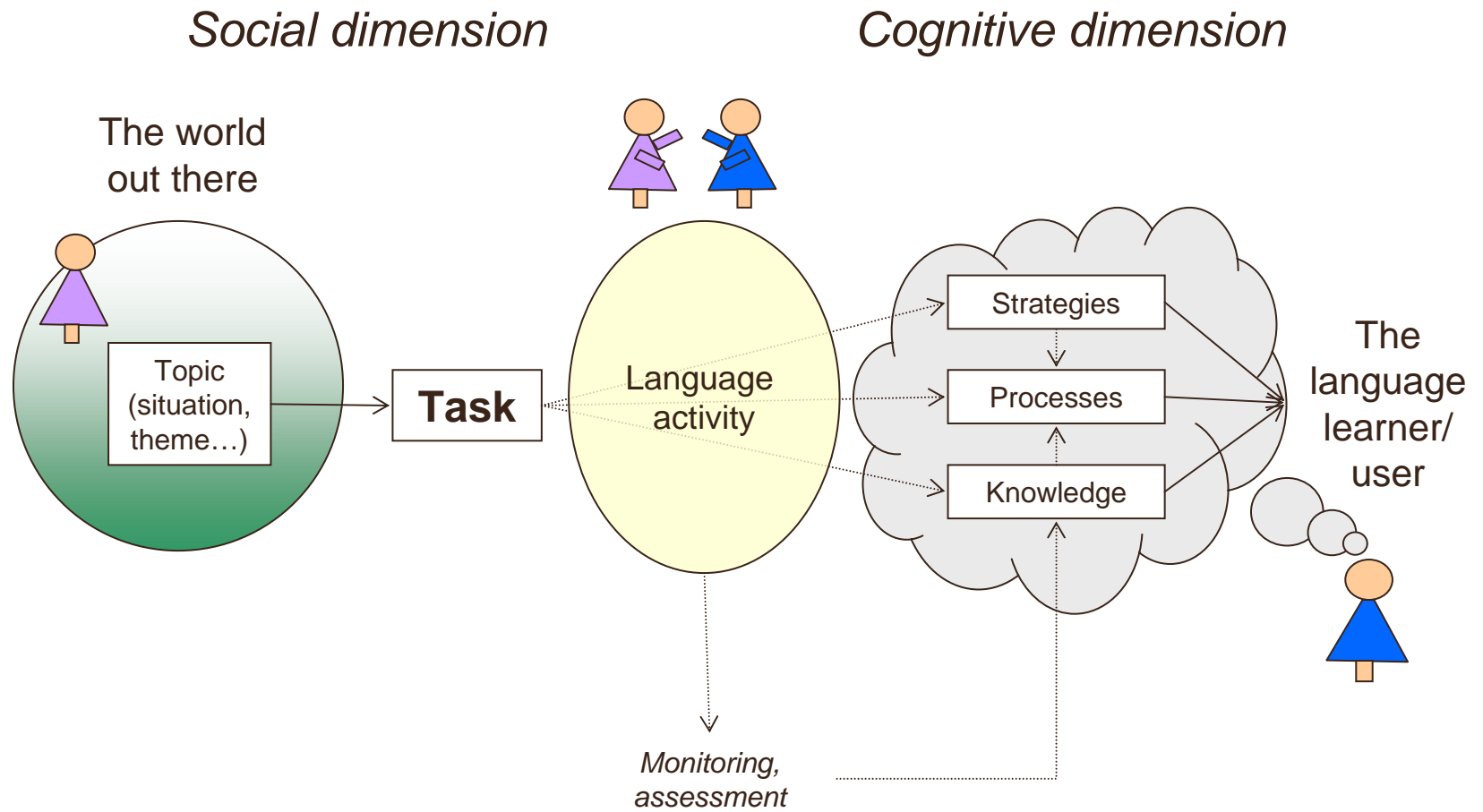


Conclusions

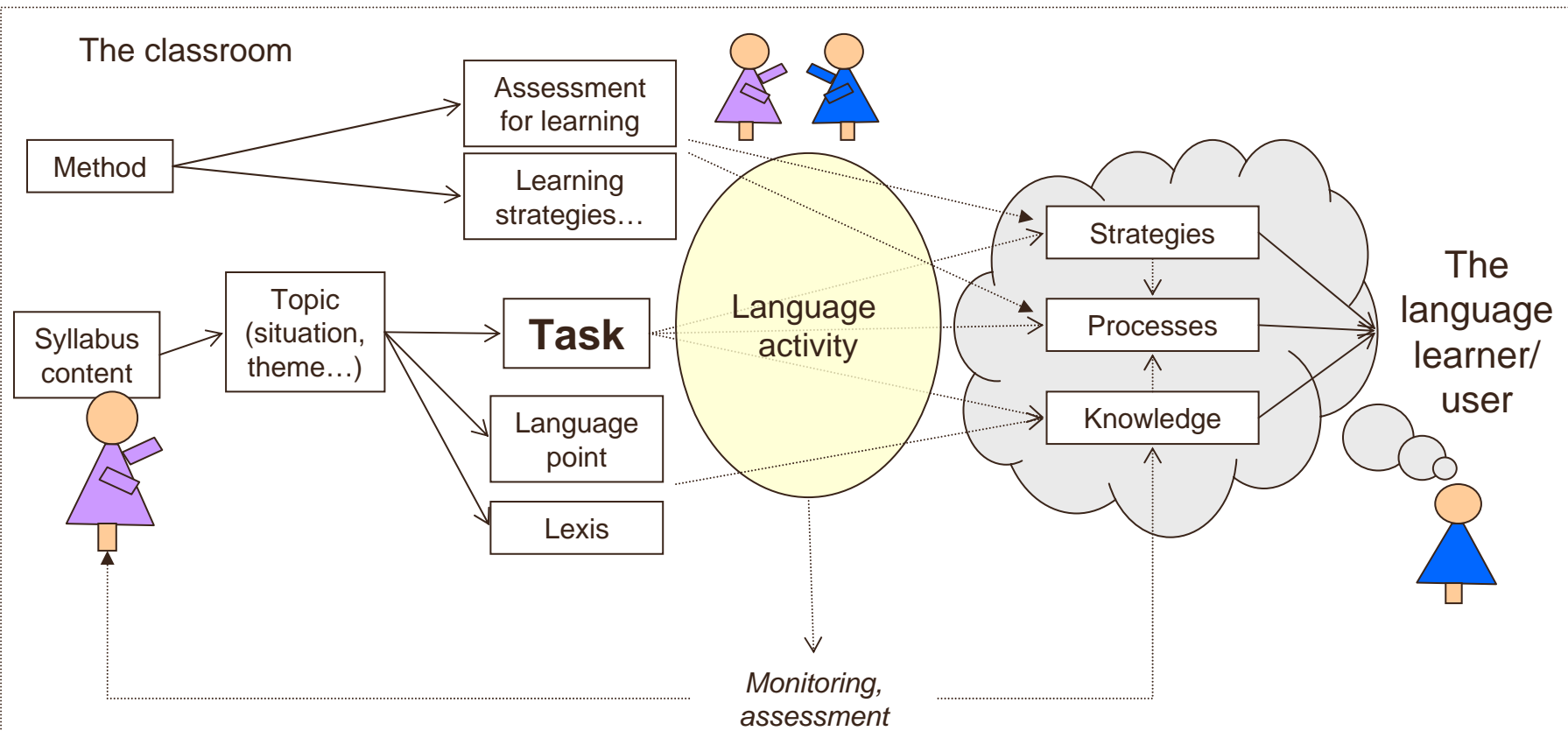
- AL illustrates difficulty of making concrete comparisons within a framework:
 - Different contexts of learning and language use (community language learners, modern foreign language learners)
 - Literacy
 - Different age groups
- Relating to a single frame of reference requires specific arguments, and realism: usefulness rather than some absolute truth.
- AL also illustrates that we work to deadlines and with finite resources: an iterative approach is needed.
- These factors impact on any claim to alignment to the CEFR. We continue to seek methods which serve the purpose.
- The Manual currently envisages users applying it creatively, “contributing to a body of knowledge and experience” and “adding to the compendium of suggested techniques”.
- We very much endorse this approach to using and perhaps further developing the Manual.



The CEFR's model of language use



Language learning in a formal setting



Cambridge ESOL Common Scale

