



UNIVERSITY of CAMBRIDGE
ESOL Examinations

Experts in Language Assessment

The classroom and the CEFR: implementing practical models of formative assessment

Neil Jones

IATEFL Testing, Evaluation and Assessment SIG and EALTA Conference
Putting the CEFR to good use
29 - 30 October, 2010





A critique of Formative Assessment

- (Randy Bennett, Seminar at Cambridge Assessment, 2010)
- The term formative assessment does not yet represent a well-defined set of artifacts or practices.
 - A meaningful definition requires a conceptual framework, theory of action, and concrete instantiation.
- Commonly made quantitative claims for the efficacy of formative assessment are suspect.
- Rooting formative assessment in pedagogical skills alone is insufficient.
- Formative assessment is assessment.
 - If it's assessment, relevant measurement principles should figure centrally in the conceptualization and instantiation.
- Formative assessment requires:
 - A Validity Argument to support the quality of inferences about students and the adjustments to their instruction.
 - An Efficacy Argument to support the impact of the inferences and adjustments.
- Each argument requires backing, both logical and empirical.



What kind of data?

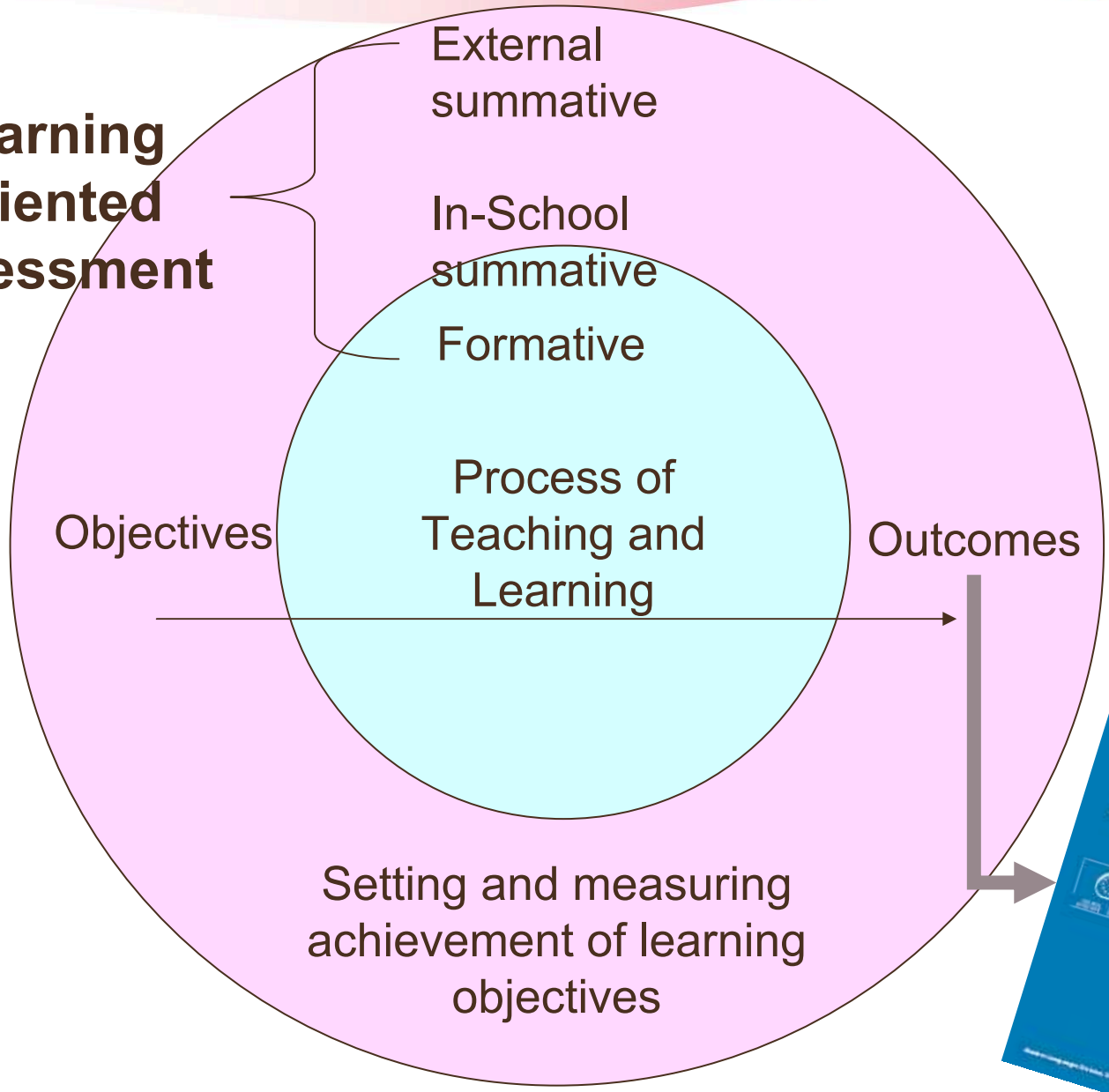
- If we accept Bennett's analysis, a key question will be what kind of data a formative system must collect in order to:
 - Work as a system – provide feedback, keep a record of progress;
 - Demonstrate the efficacy and validity of the system.
- Not just classroom tests: a range of structured interactions/exercises that produce recordable evidence
- Not limited to items of course content – other learning skills and strategies
- Emerging from learners' reflection and self-assessment
- Applicable both to the case of the teacher-led classroom and the machine-led learning environment, or to some blended combination of these.
- Certainly data on the performance of individual learners would seem to require significant systems support.



Learning-oriented assessment (LOA)

- The term *formative assessment* has been so much abused that some people refuse to use it at all.
- It also suggests an opposition with *summative assessment* which is not helpful.
- Cambridge ESOL's traditional exams are summative, but have always been seen as part of a learning process
- (a process which is well supported by Cambridge ESOL and by other providers in language education).
- Hence the alternative proposal: Learning-oriented assessment (© Purpura 2008)

Learning Oriented Assessment




CEFR

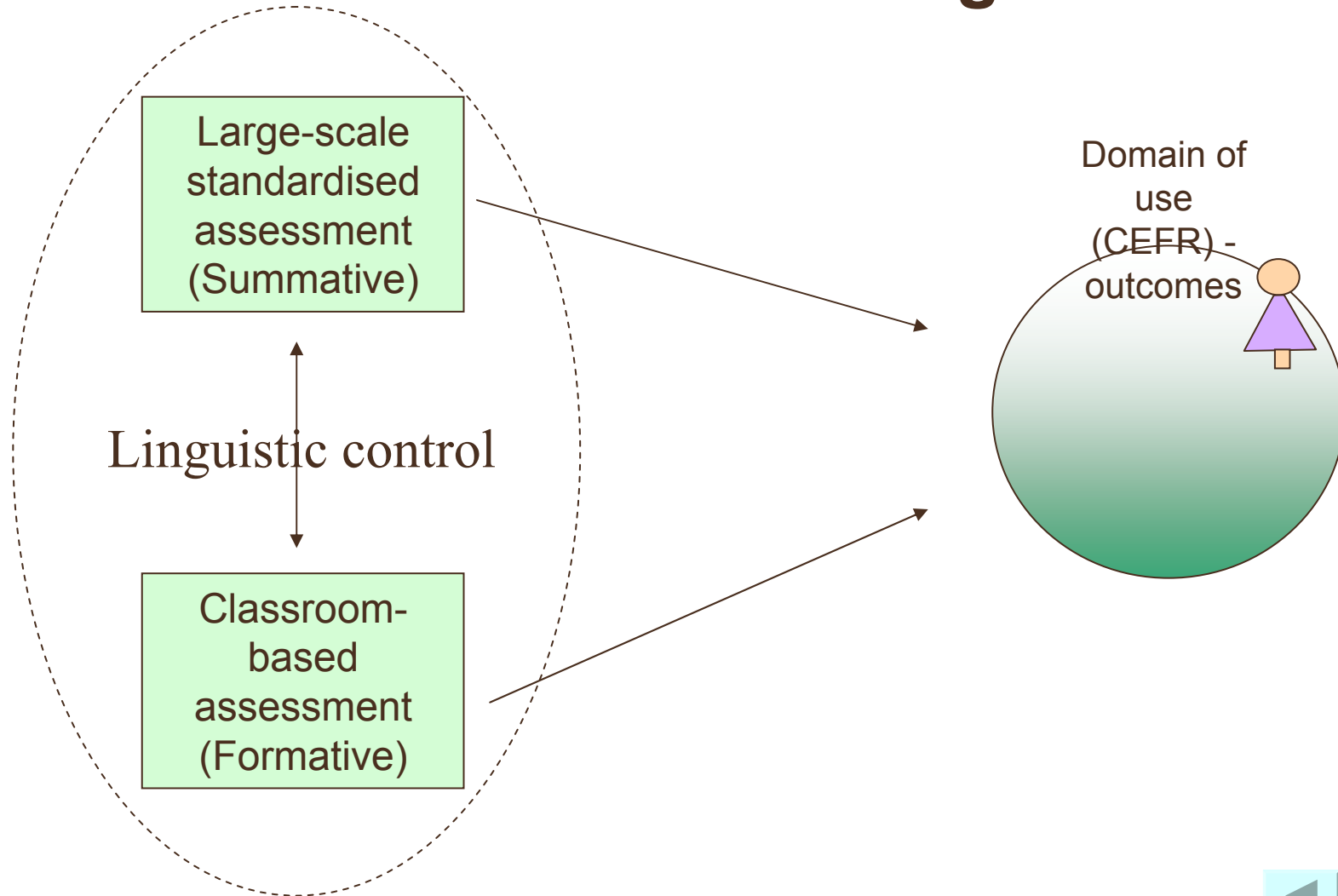




Why link classrooms to the CEFR?

- To emphasize useful outcomes, motivate, empower learner autonomy
- For an assessment body, to construct a single frame of reference for summative and formative assessment, relating to the same targeted outcomes. 
- Given the prominence of the CEFR, to address the validity argument for relating contexts of learning to it. LOA provides the methods for building such an argument.

Classroom work and learning outcomes

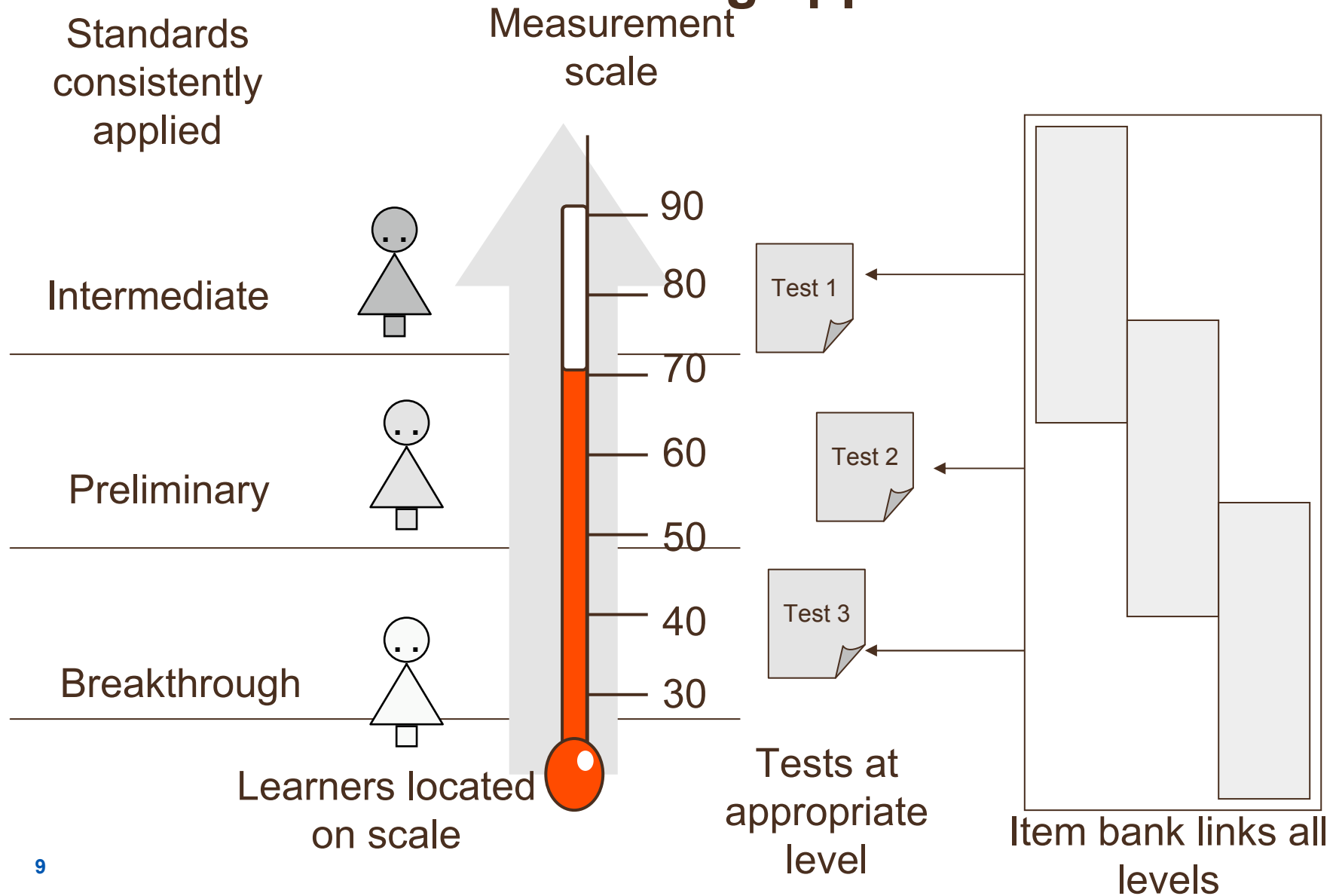




Summative and formative: how much is measurable?

- Cambridge ESOL's (objective) tests are based on an item banking approach exploiting *Latent Trait Theory*.
- “Latent trait” is the language of behavioural psychology:
- A *trait* is an enduring, quantifiable “something” that a learner has in his head.
- Even if we don't believe that language ability is like this, trait-based measurement is still very useful, because we find that even very complex phenomena (such as language ability) can be described well by simple scales.
- That is, as long as we are prepared to define what we mean by *language ability* quite specifically.

Measuring proficiency: An item-banking approach





The measurement metaphor

- We borrow the idea of measurement from the physical sciences.
- We can measure physical properties of objects using simple numerical scales:
 - For length, e.g. in centimetres
 - For temperature, e.g. in degrees celsius.
- A single number captures everything perfectly.
- This is *not* true when we measure language proficiency.
- The metaphor is useful – but imperfect.



Why can't language proficiency be measured perfectly?

- A) Because language tests aren't as precise as rulers or thermometers.
- B) Because language proficiency is more complex than length or temperature.
- The right answer is ...
- B!
- Different measuring instruments (tests) test different things – language tests are not interchangeable.
- Even two learners placed at *A2* by the *same* test will have distinct differences.



Narrowing the focus

- Learners differ in many ways:
- Profile of skills: Reading, Writing, Listening, Speaking
- (But we can deal with this by testing each skill separately).
- Age, purpose for learning, domain of use...
- (But we can deal with this by designing tests for specific well-defined groups of learners).
- These kind of differences can be handled well within the measurement metaphor.
- But there is a price to pay: these different tests are not directly comparable.
- And there's a limit to how far you can go.



The limits of measurement

- Differences that are difficult to deal with within the measurement metaphor:
- Individual learning histories,
- conscious / unconscious knowledge of language,
- learning styles,
- use of strategies,
- motivation, confidence, independence, etc.
- These are differences between individual learners: qualitative rather than quantitative.
- Accommodating qualitative differences is a significant aspect of formative assessment.



Quantitative and qualitative aspects are complementary

- Trait-based proficiency measures:
- Provide goals, personal objectives, motivation
- Provide an orientation in progress made
- Help individualise learning: for each learner there is a level which is optimal for learning.
- Enable teaching to focus on individual strengths and weaknesses, things which are helping or hindering the learner.
- Quantitative measures facilitate qualitative diagnosis and action.



Two models of learning

Adaptivity

- Learning happens by exposure to comprehensible input.
- Formal language learning (teaching) does not work
- Chomskyan: Language acquisition is an innate capacity
- Comprehensible input is at a level just beyond the learner's current capacity to use (the $i+1$ level)
- (Focus on the quantitative: Item banking enables us to implement the $i+1$ level).

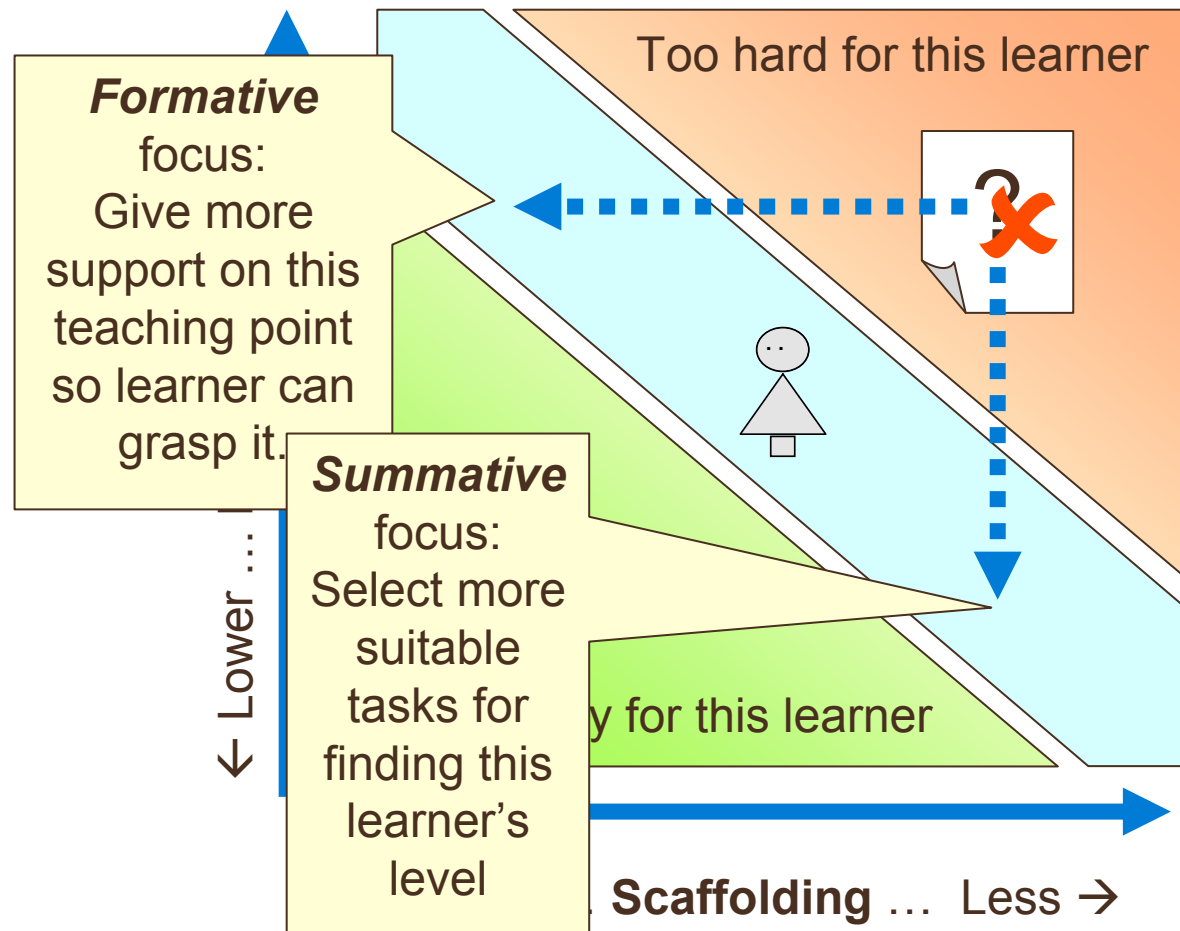
Krashen

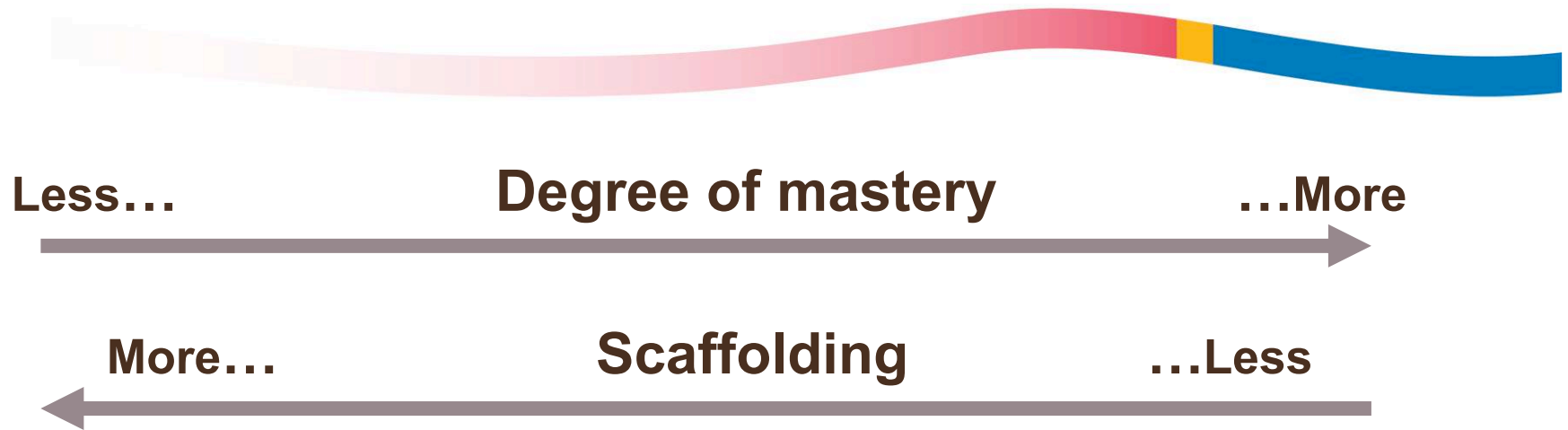
Interaction

- Learning happens through interaction with a more knowing other (e.g. a teacher).
- i.e. all cognition is socially constructed.
- What a learner can achieve with assistance defines the *zone of proximal development*.
- (Focus on the qualitative: These ideas have been influential e.g. on Dynamic Assessment, Assessment for Learning..).

Vygotsky

Zone of proximal development

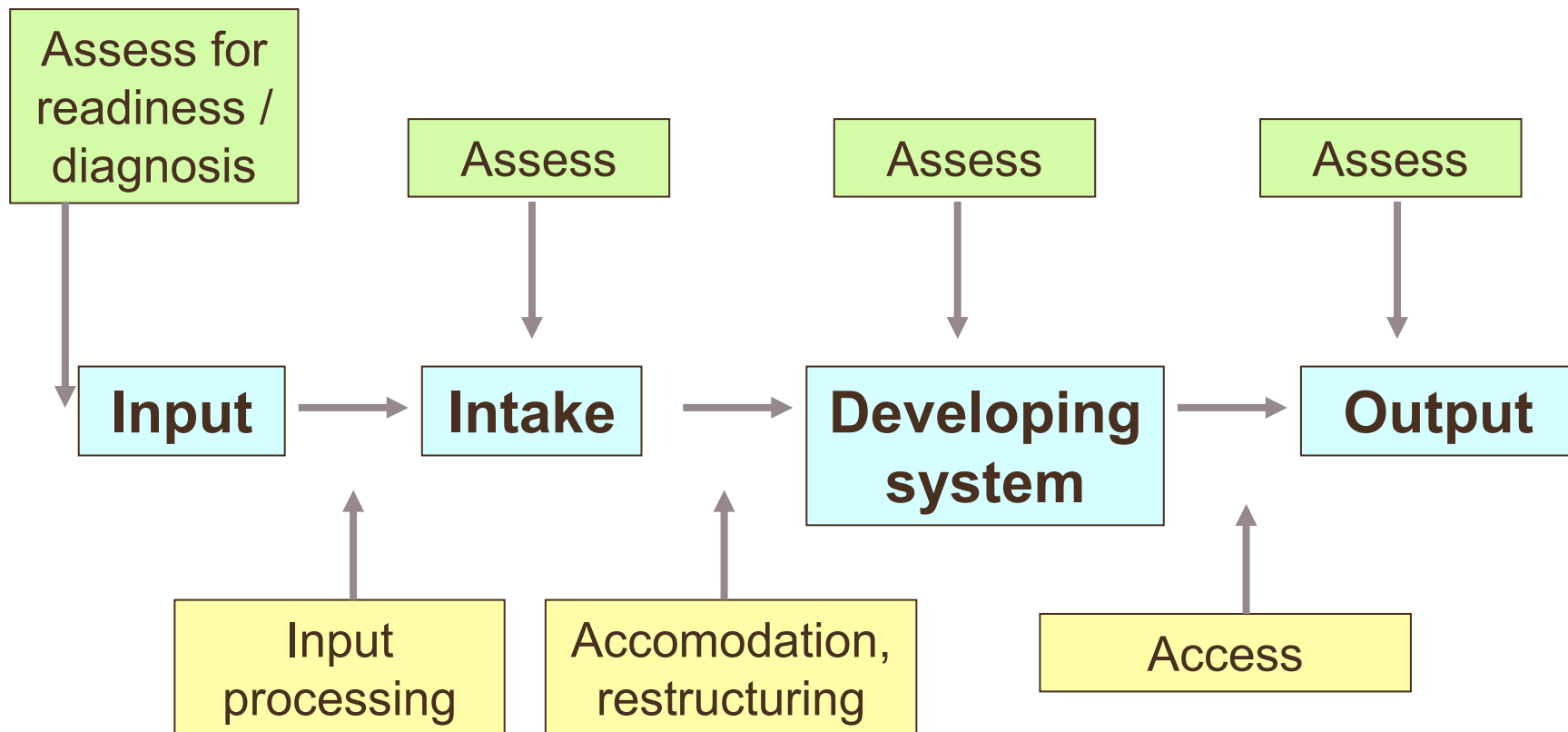




- Scaffolding compensates for differential degrees of mastery.
- This is what makes classroom performance difficult to interpret.
- Degrees of mastery are reflected at a number of levels:
- Standard teacher routines: *Presentation* → *drill* → *practice*
- Skills: *Receptive* → *productive*
- Tasks: *Context embedded, concrete* → *context reduced, abstract*
- Focus on form: *Conscious, careful* → *automatic, fluent*
- Cognitive learning model: *Input* → *intake* → *developing system* → *output* (Van Patten)

Testable moments

(adapted from Purpura 2008)

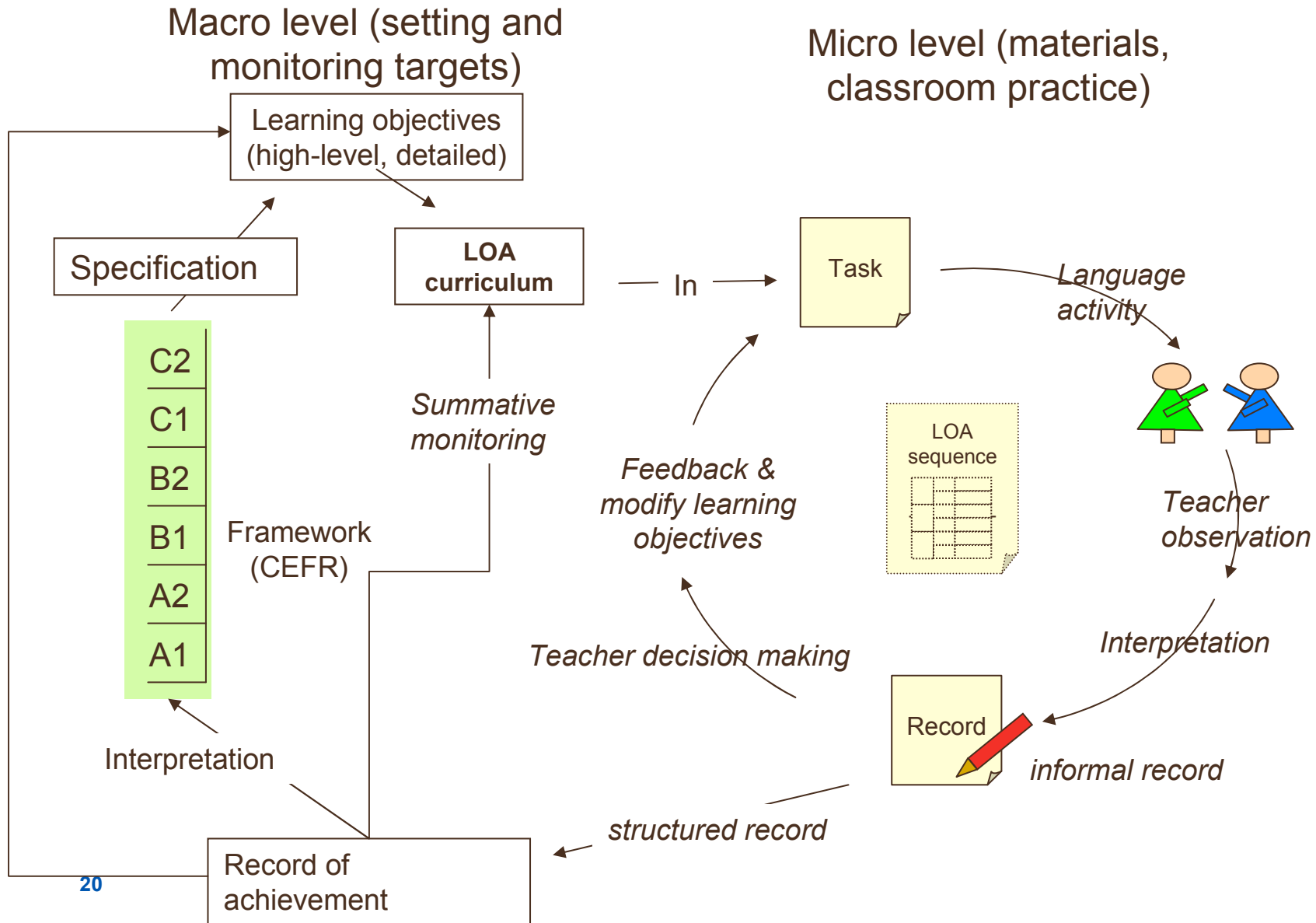




Can-dos for the classroom

- “Testable moments” should not imply a reductive conception of language learning and teaching.
- Can-dos for the classroom include:
- Curricular objectives, course content
- Learning how to learn, *assessment for learning* objectives
- Language awareness, knowledge about language
- Cultural issues, inter-cultural awareness, affect
- i.e. *language is unique* in terms of the different levels at which it engages the successful learner

Testable moments: a general LOA scheme

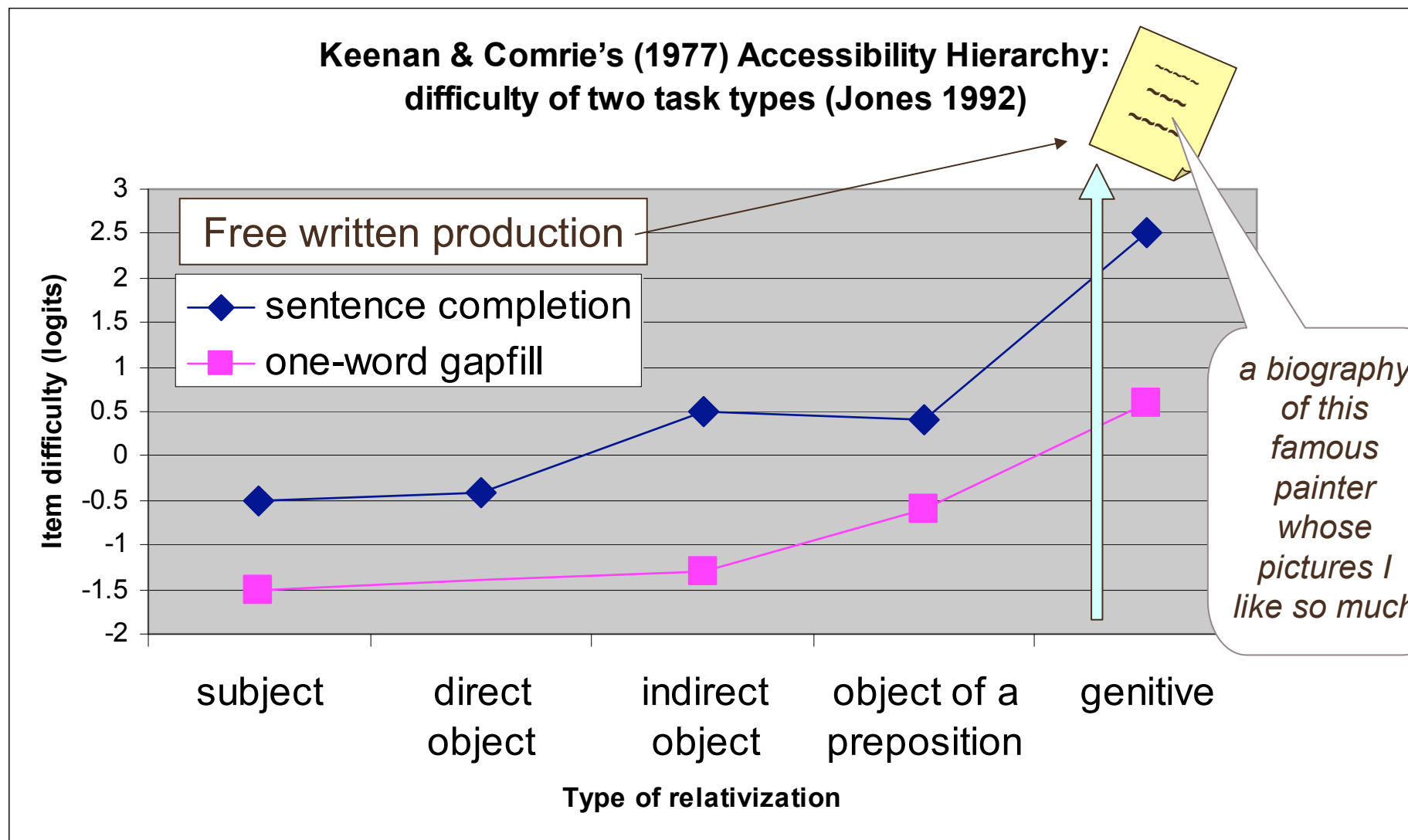




testable moments: the English Profile and criterial features

- The EP is finding evidence of structures which appear in Writing in the CLC at certain CEFR levels, and are thus *criterial features* of the level.
- E.g. genitive relatives (*... a biography of this famous painter whose pictures I like so much*) are first found at B1.
- This kind of information might inform when or in what order to teach different kinds of relativisation.
- Presumably such structures are attended to and become part of a developing system *before* they appear in free output.
- Such partial mastery could be tested for using different forms of elicitation.

Eliciting degrees of mastery of structure





How to interpret classroom performance in criterion-related terms?

- In-school or external summative assessment are a practical way of linking to high-level objectives.
- That leaves the problem of how to link formative classroom assessment to the summative assessment.
- Formative interpretations might take the form: learner x is at point y on the way to fully mastering objective z . i.e. factor in degrees of mastery.
- The models illustrated above might enable us to do this.
- Interpretation thus involves a forward-looking prediction of how close the learner is to achieving an objective, based on either:
 - observed average learning gains; or
 - The individual learner's current trajectory.



Can assessment really look forwards?

- Summative assessment is backward-looking (it measures past achievements).
- Formative assessment is said to be forward-looking in the sense that it aims to promote future learning.
- Dynamic Assessment claims additionally that a learner's observed capacity to benefit from formative interactions is a better predictor of future learning than summative assessment.
- This suggests that an ongoing record of formative interactions would enable us to estimate learning *trajectories*, not just learning achievements to date.
- Quite a challenge for the formative assessment enterprise.