Measuring second language development: Issues and challenges in instructed SLA research

Andrea Révész

UCL
Institute of Education

Introduction

➤ Primary research area: second language acquisition (SLA)

➤ BUT keen interest in assessment

➤ Interested in identifying interfaces between SLA and assessment

➤ Aim to incorporate insights from assessment into SLA research

Aim of this talk …

➢ Discuss some methodological issues and challenges in measuring second language development in the context of instructed SLA research.

➢ Focus on issues in operationalising and measuring various types of L2 knowledge.
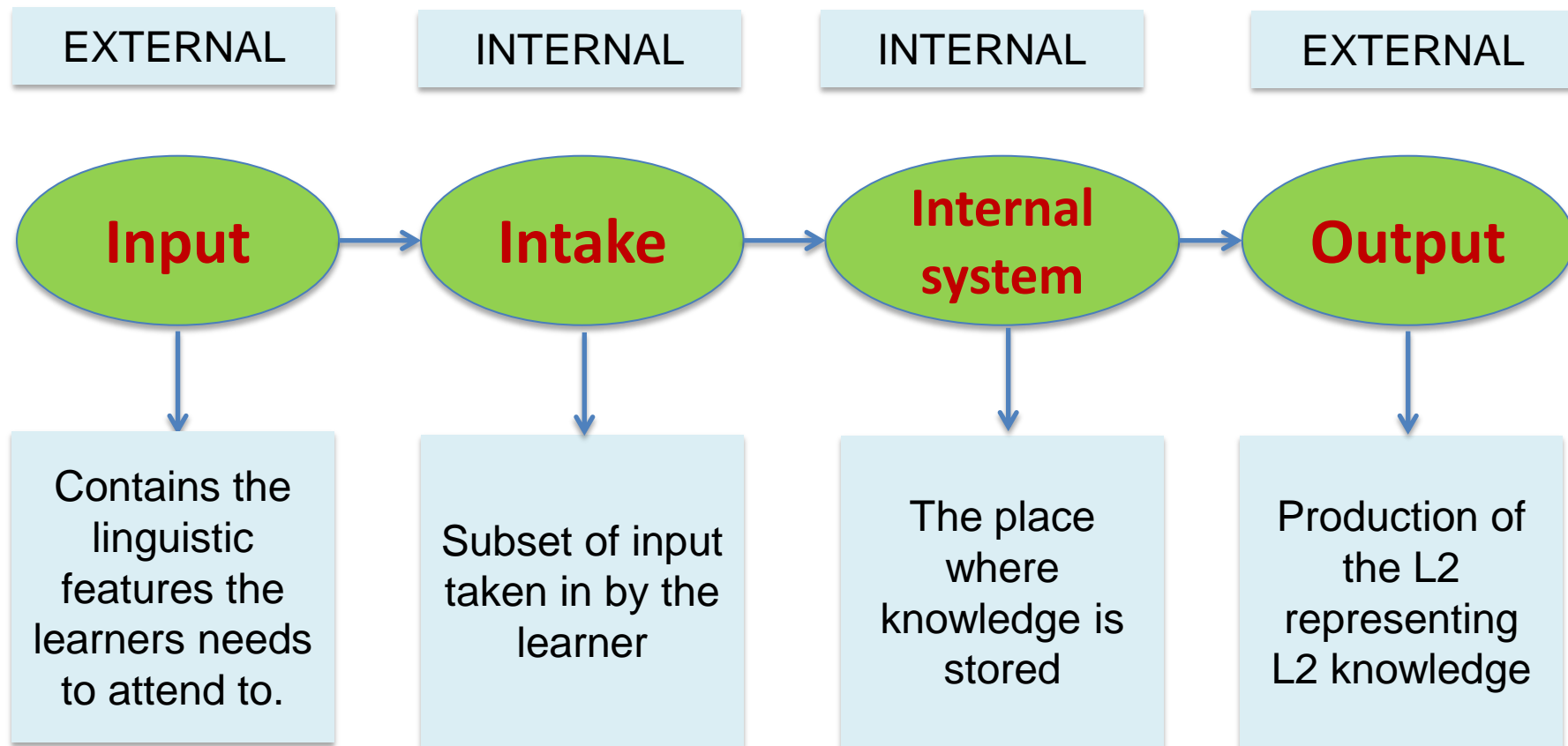
What is instructed SLA?

Instructed second language acquisition (SLA) research investigates formal second language (L2) learning and has as its ultimate goal the understanding and improvement of instructional practices within educational settings.
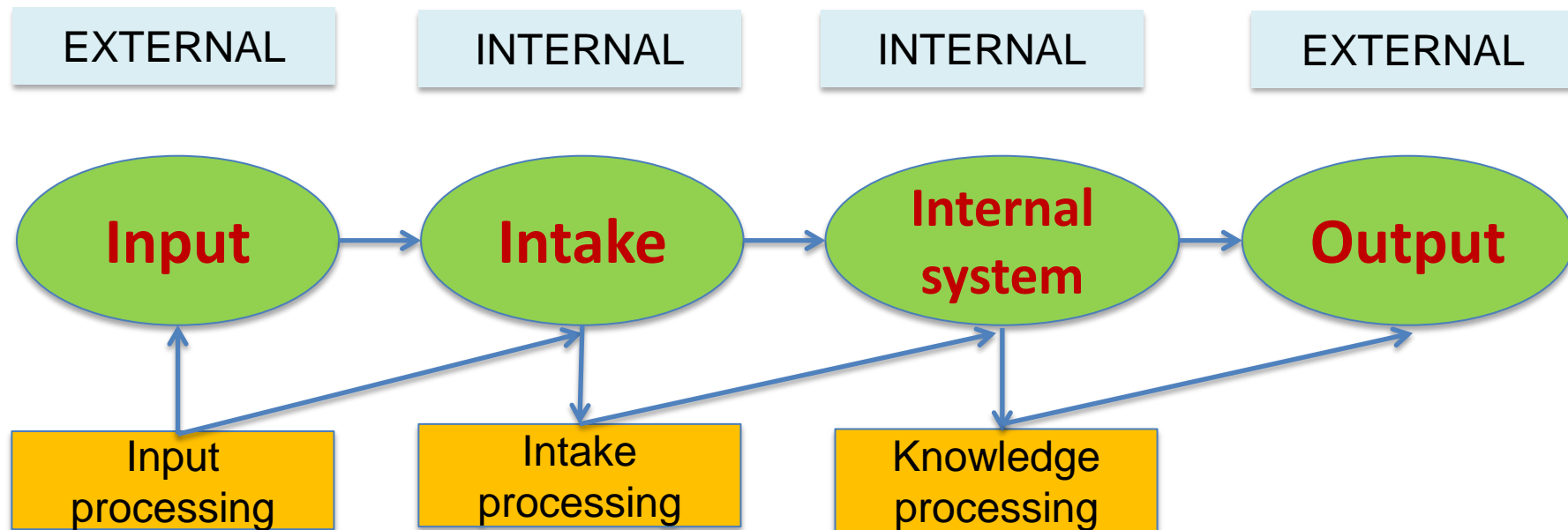
(Ortega, 2005, p. 319)

A simple model of SLA
(Gass, 1997; Schmidt, 2001; Leow, 2015)

EXTERNAL

INTERNAL

INTERNAL

EXTERNAL

Input → **Intake** → **Internal system** → **Output**

Contains the linguistic features the learners needs to attend to.

Subset of input taken in by the learner

The place where knowledge is stored

Production of the L2 representing L2 knowledge

A simple model of SLA
(Gass, 1997; Schmidt, 2001; Leow, 2015)

EXTERNAL   INTERNAL   INTERNAL   EXTERNAL

Input → **Intake** → **Internal system** → **Output**

Input processing   Intake processing   Knowledge processing

A key aim of cognitive instructed SLA research is to identify ways in which we can facilitate input, intake, and knowledge processing to promote L2 development in instructed settings.

How do researchers go about achieving this goal?

➢ Many SLA studies are interventionist and developmental

| PRE-TEST | TREATMENT | POST-TEST | DEL. POST-TEST |

➢ Aims to establish the learner's existing or current knowledge of the targeted linguistic construction.

➢ Focus is on a single or a narrow range of linguistic features whose knowledge the treatment is designed to facilitate (e.g., article system, past tense, vocabulary items)

How do researchers go about achieving this goal?

➢ Many SLA studies are interventionist and developmental

| PRE-TEST | TREATMENT | POST-TEST | DEL. POST-TEST |

➢ Can take a large variety of forms. For example,
   ➢ oral/written corrective feedback
   ➢ textual enhancement
   ➢ task-based work

➢ The aim is to facilitate the input, intake and knowledge processing stages of the developmental process.

How do researchers go about achieving this goal?

➢ Many SLA studies are interventionist and developmental

| PRE-TEST | TREATMENT | POST-TEST | DEL. POST-TEST |

➢ Typically takes an equivalent form of the pre-test.

➢ It aims to test whether any changes in knowledge have occurred due to the treatment.

How do researchers go about achieving this goal?

PRE-TEST	TREATMENT	POST-TEST	DEL. POST-TEST

In the remainder of this talk, I discuss some methodological issues and challenges in measuring **knowledge at the pre-test and post-test stages**, with a view to the potential **threats of construct underrepresentation**.

What is exactly tested at the pre-test and post-test stages?

What constructs are typically tested at the pre-test/post-test/delayed post-test stage?

The aim of the pre-test and post-test is to establish the learner's existing or current knowledge of the targeted linguistic construction.

How is knowledge conceptualised in cognitive approaches to instructed SLA?

Second language knowledge

Declarative/Explicit	Procedural/Implicit
Knowledge that	Knowledge how
Can be verbalised	Can be performed
Slow	Efficient
Generalisable	Specific, hard to transfer

The aim of instructed SLA is to **foster development in both of these knowledge types**, even if the ultimate goal is the development of procedural knowledge or the ability to use the L2.

How can we measure explicit and implicit knowledge? (e.g., Ellis et al., 2009)

Declarative/Explicit	Procedural/Implicit
No time pressure	Time pressure
Focus on form	Focus on meaning
Metalinguistic knowledge involved	Metalinguistic knowledge not involved
Response according to rule	Response according to feel
➢ untimed grammaticality judgment task (GJT) ➢ multiple-choice tests ➢ written production tests	➢ elicited imitation ➢ spontaneous production ➢ timed grammaticality judgment task (GJT) test

Digital repository data collection instruments
for research into second language learning and teaching

Threats to construct underrepresentation in measuring L2 knowledge

➢ Failure to use multiple measures of development tapping various types of L2 knowledge

➢ Lack of match between treatment and assessment of knowledge

➢ Using insensitive measurement units to assess development

➢ Adopting a purely product-oriented perspective of L2 knowledge and development

Need to elicit multiple indicators of development

➢ Most instructed SLA studies tend to rely on a limited number of tests, often assessing the same type of knowledge.

  ➢ For example, in studies of recasts, a single type of oral communication task is typically used to tap changes in learners' interlanguage (e.g., Long et al., 1998; Mackey & Philp, 1998; Révész, 2009).

  ➢ Few studies incorporate multiple outcome measures gathered from various linguistic and communicative contexts (Ammar & Spada, 2006; Ellis et al., 2006; Han, 2002; Révész, Sachs, & Hama, 2014).

Need to elicit multiple indicators of development

Why is it important to assess development using a variety of measures?

Lack of multiple behavioural observations can introduce a missing variable bias, since a form or structure may emerge in different contexts.

The potentially different effects of instruction on distinct types of linguistic knowledge may be overlooked unless studies include a variety of carefully selected measures.

Need to elicit multiple indicators of development

➢ Lack of multiple data sources may result in interpretations that are "based on a lack of evidence, as opposed to evidence for the lack of emergence" (Norris and Ortega, 2003, p. 733).

➢ "At the very least, both types of measures, implicit and explicit, should be employed" in instructed SLA research. (Doughty, 2004, p. 193).

➢ Ellis (2009) and colleagues also call for research employing valid measures of both implicit and explicit knowledge.

Révész (2012) …

… was among the first studies to use multiple outcome measures to assess the effects of an instructional intervention on second language development.

Révész (2012)

Group	Pretest 1st day	Session1. 1st day	Session1. 2nd day	Session1. 3rd day	Posttest 4th day	Del. Post. 4 weeks later
Recast	Pretest	Recast	Recast	Recast	Posttest	Del. Posttest
Non-recast	Pretest	Non-recast	Non-recast	Non-recast	Posttest	Del. Posttest
Control	Pretest				Posttest	Del. Posttest

LAST SUNDAY, 1:00 pm

Recasts

Correct reformulations of the learners' incorrectly formed utterance.

Learner utterance:

And I saw a boy next to the bar. I think he was with his girlfriend. ***They talking to each other.***

Recast:

They ***were talking*** to each other.

Pretest/Posttest/Delayed Posttest

➢ **Grammaticality Judgment Task**

  ➢ Designed to test explicit knowledge

➢ **Written Picture Description Task**

  ➢ Expected to elicit both implicit and explicit knowledge

➢ **Two Oral Photo Description Tasks**

  ➢ Designed to assess implicit knowledge

Results of many-facet Rasch measurement revealed …

	Gram Judgment Task	Written Description Task	Oral Description Tasks
Recasts > Nonrecast	✔	✔ ✔	✔ ✔ ✔

Relying on a single measure would have yielded a partial picture of the effects of instruction.

Threats to construct underrepresentation in measuring L2 knowledge

➤ Failure to use multiple measures of development tapping various types of L2 knowledge

➤ Lack of match between treatment and assessment of knowledge

➤ Using insensitive measurement units to assess development

➤ Adopting a purely product-oriented perspective of L2 knowledge and development

Matching treatment and assessment

Another potential shortcoming in instructed SLA studies is a lack of correspondence between the outcome measures and the instructional treatment (e.g., Norris & Ortega, 2003).

.

For example,

	Gram Judgment Task	Written Description Task	Oral Description Tasks
Recasts > Nonrecast	Little match	Better match	Good match

Rationale for matching treatment with assessment

➢ Transfer-appropriate processing:

  ➢ We can better transfer and "remember what we have learned if the cognitive processes that are active during learning are similar to those that are active during retrieval" (Lightbown, 2007, p. 27)

➢ Implication of transfer-appropriate processing
  ➢ "learning to use language in a communicative context may improve the ability to retrieve it in such contexts" (Lightbown, 2007, p. 27)

Matching treatment with assessment

➢ Instructed SLA research does not fare well when it comes to applying theTransfer-appropriate processing principle

➢ Norris & Ortega's (2000) meta-analysis

  ➢ Regardless of treatment type, SLA researchers tend to use pretest-posttest measures that are explicit (e.g., multiple-choice tests, cloze tasks)

  ➢ Thus, there is a bias in favour of more explicit interventions as opposed to more implicit treatments

➢ A growing number of studies use measures that reflect the instructional treatment.

Threats to construct underrepresentation in measuring L2 knowledge

- Failure to use multiple measures of development tapping various types of L2 knowledge

- Lack of match between treatment and assessment of knowledge

- Using insensitive measurement units to assess development

- Adopting a purely product-oriented perspective of L2 knowledge and development

Using interlanguage sensitive measures of development

➢ The validity of outcome measures may be compromised unless appropriate measurement units are employed to assess learners' performance (e.g., Norris & Ortega, 2003).

➢ Changes in accuracy typically used as criterion for development

  ➢ Accuracy scores are often computed for extended oral and written production tasks by calculating percentage of accurate use in obligatory contexts for a certain construction.

However, accuracy is a conservative criterion for establishing development …

➢ Emergence (i.e., first productive use) is considered a developmental milestone

➢ Measures of accuracy cannot account for non-linear development such as U-shaped behaviour

➢ Target-like accuracy ignores evidence of changes that may show improvement while still not reaching the full target variant

Using interlanguage sensitive measures of development

How should we define measurement units in SLA research then?

Adopt analytical frameworks in which emergence, measures of partial mastery, and target-like accuracy are all combined (e.g., Norris & Ortega, 2003).

Multiple assessments? Complement target-like scores with partial scoring (e.g., 2 points correct, 1 point partially correct)?

In a similar vein, Purpura (2005) recommended …

… employing partial scoring in multiple choice tests depending on the extent to which the distractors approximate the target-like variant.

I _____ to the swimming pool yesterday.

A. Go   0 points

B. Goed   1 point

C. Wented   1 point

D. Went   2 points

Threats to construct underrepresentation in measuring L2 knowledge

- ➢ Failure to use multiple measures of development tapping various types of L2 knowledge

- ➢ Lack of match between treatment and assessment of knowledge

- ➢ Using insensitive measurement units to assess development

- ➢ Adopting a purely product-oriented perspective of L2 knowledge and development

Looking at development in both process and product

➢ Existing SLA research tends to operationalise L2 development adopting a **product-oriented view**.

➢ It is also possible, however, to view L2 development from a more **processing-oriented perspective**, in other words, to see whether there are **changes in terms of how and how fast learners process** L2 features due to an instructional treatment they received.

Looking at development in both process and product

Quantitative differences in processing

Measuring speed of processing via reaction times

- ➢ A sign of L2 development could be a decrease in speed or reaction times even if the accuracy rate remains stable.

- ➢ So far reaction times have rarely been used to document development (BUT, when we use language, speed is often as important as accuracy).

Grammaticality Judgment Task

Yesterday I go to Reading.

Correct   Incorrect
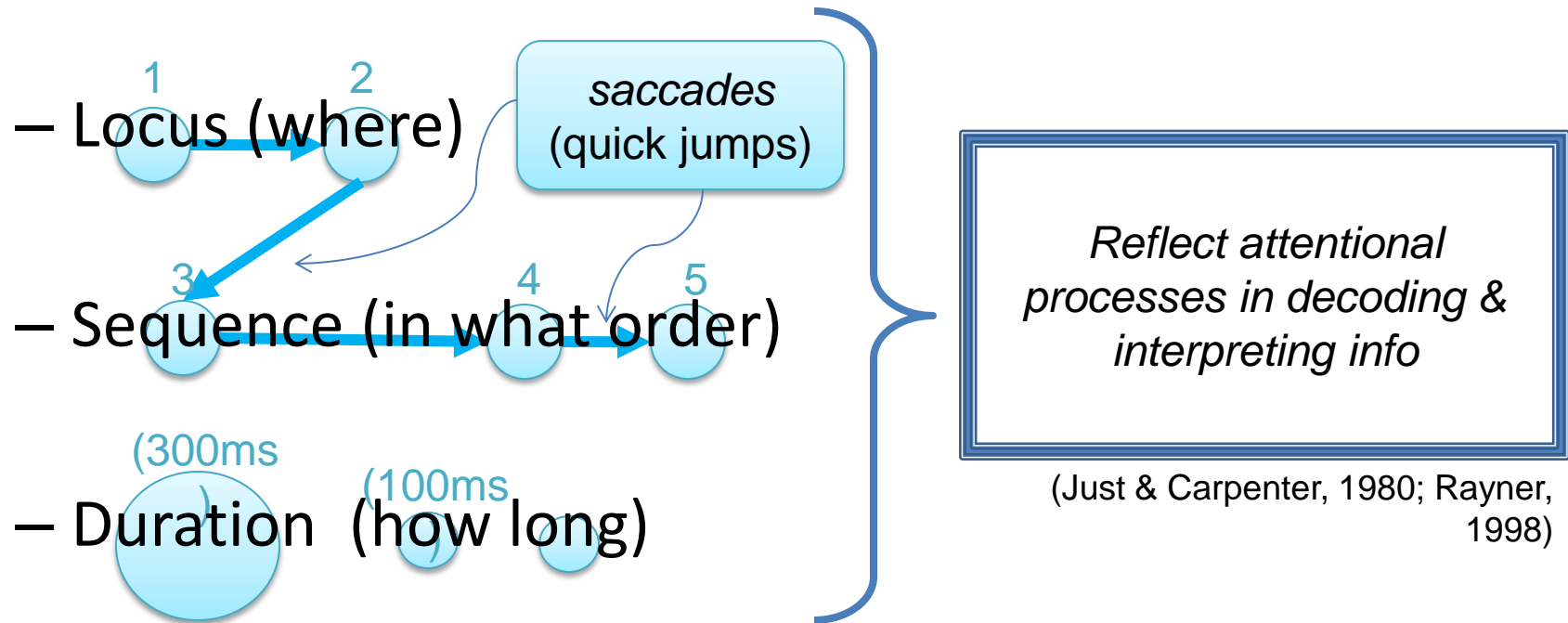
Yesterday I go to Reading.

Looking at development in both process and product

Qualitative differences in processing

Eye-tracking might offer one way to identify qualitative changes in processing written language.

Eye Tracking

Recording the moment-by-moment **eye fixations** of an individual interacting with a visual stimulus

- Locus (where)

saccades (quick jumps)

- Sequence (in what order)

- Duration (how long)

Reflect attentional processes in decoding & interpreting info

(Just & Carpenter, 1980; Rayner, 1998)

1 2 3 4 5

(300ms (100ms

Sagarra and Ellis (2013)

➢ They believe that **yesterday** the boy **cooked** something for the party.

➢ They believe that **yesterday** the boy **cooks** something for the party.

➢ They believe that the boy **cooked** something **yesterday** for the party.

➢ They believe that the boy **cooks** something **yesterday** for the party.

To summarise,

Discussed some methodological issues and challenges in measuring second language development, with a view to decreasing the risk of construct underrepresentation:

➢ Highlighted the importance of utilising multiple outcome measures

➢ Drawing on the notion of transfer-appropriate processing, I also emphasises the need for including assessment instruments that closely match the instructional treatment.

➢ Emphasised the importance of selecting interlanguage sensitive measurement units that capture partial mastery of L2 forms.

➢ Argued that it would be important to move away from a purely product-centred conceptualisation of L2 development to a more processing-oriented perspective.

Thank you!