



AI for dynamic and diagnostic assessment: Automatic task design and mediation to support development of L2 English reading

Ari Huhta
Dmitri Leontjev
Roman Yangarber
Matthew Poehner

July 3

45th Language Testing Research Colloquium, Innsbruck

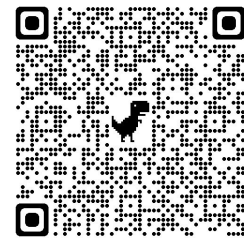


Our aim: to address two issues in assessment that supports L2 learning

- **Lack of systematicity** and **basis in theory** of learning / development in the skill(s) of interest, particularly in formative assessment but also:
 - **Diagnostic assessment**: lack of specificity in how to use assessment results to support learning/teaching
 - **Dynamic assessment**: lack specificity/systematicity in construct identification
- **Lack of individualisation** due to large groups and limited teacher time
 - potential issues for the quality and impact of assessment
 - possible consequences include learners not receiving instruction that optimally supports their language development



What do we do to address these issues?



We conduct research that aims to advance the **theoretical, empirical and practical basis** of L2 assessment for learning

DD-LANG research project: Dynamic-diagnostic language assessment – a conceptual and practical innovation in foreign language assessment

- funded by the **Research Council of Finland** and University of Jyväskylä Sept.1. 2022 – Aug. 31. 2026
- focuses on reading and writing in English as a foreign language in upper secondary schools in Finland (16-18 -year-olds)



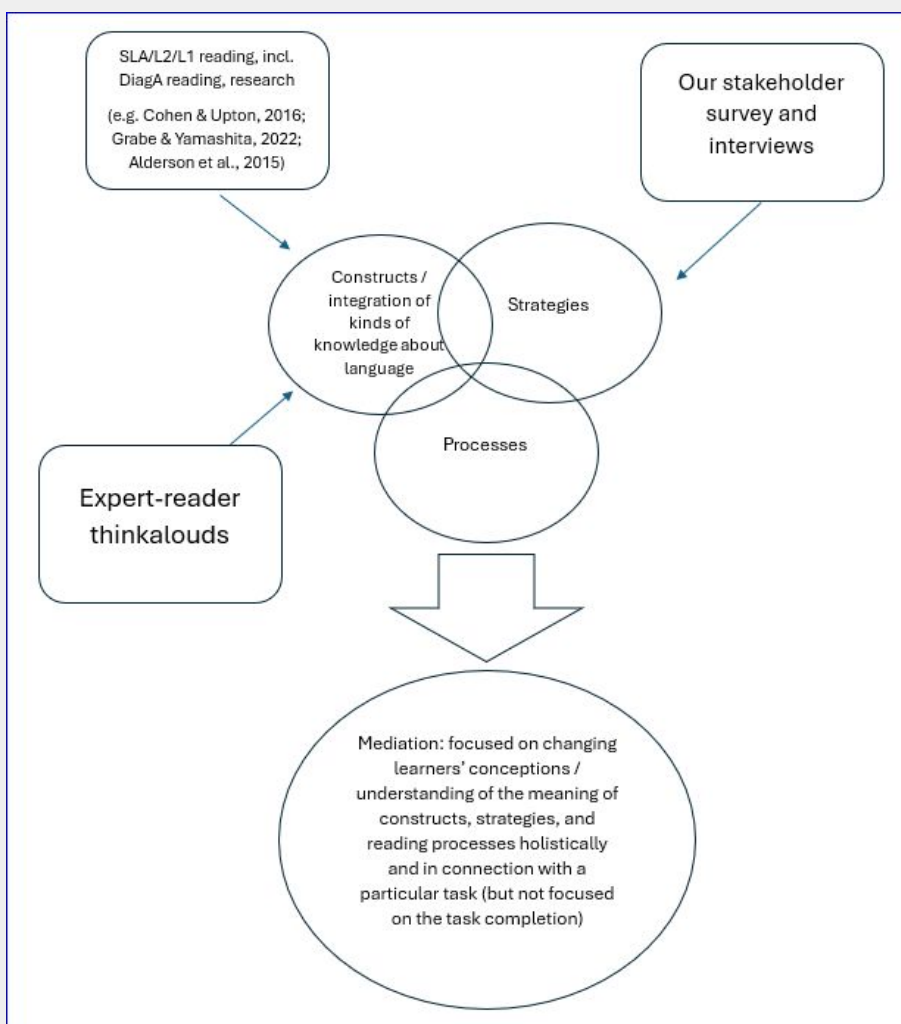
How do we address these issues in DD-LANG?



- (1) By integrating L2 dynamic assessment (DA) and diagnostic assessment (DiagA)
 - clearer theoretical underpinnings of (a) **assessment constructs** (DiagA) and (b) **development** of learner L2 abilities (DA)
- (2) By increasing synergy of teacher-led classroom assessment & AI-enhanced online assessment in Revita
 - provides teachers and learners with (a) more, and more individualised, **information about learner development** and (b) **input to classroom activities**



Defining and operationalising reading constructs





Two types of **constructs** & Two assessment **contexts**

- **Narrower constructs**

- e.g., discourse markers, verb tenses, articles, ...
- online tasks created automatically by using NLP / AI
- mediation created mostly by researchers
- **context:** mainly online

- **Broader (reading) constructs**

- e.g., inferring word meanings from context, understanding main idea
- tasks & mediation for each construct created by researchers
- **context:** online and in the classroom
- AI: rephrase and translates mediation



Narrower construct example: discourse markers

Deer in the Snow

A snowflake landed right on the tip of his left thumb. It seemed to look back at him as it dissolved. Behind it, in the distance, there was fluttering. He lifted his binoculars. It was a deer, walking towards him. How much closer would it get before it picked up his scent? He was fourteen feet up, and the wind was to his face, he could afford to be patient.

in short

so

then

when

Check Answers

Next snippet ↓

Report a problem



Narrower construct example: discourse markers

Deer in the Snow

A snowflake landed right on the tip of his left thumb. It seemed to look back at him as he lifted his binoculars. It was a deer, walking towards him. How much closer would it get if the wind was to his face, then he could afford to be patient.

Check Answers

Next snippet ↓

- Read around the gap.
What ideas can you find in the text surrounding it? How does the missing part connect these ideas? Think how you as a writer help your reader to understand connections between your ideas.
- Words like this can show different relationships between ideas they connect.
These words can show emphasis or time sequence.
How do you think the ideas are connected here?
- Look at the following example; think which of the options expresses the same relationship: The weather was hot.
Hence, my ice-cream melted.



Narrower construct example: discourse markers

Deer in the Snow

A snowflake landed right on the tip of his left thumb. It seemed to look back at him as it dissolved. Behind it, in the distance, he lifted his binoculars. It was a deer, walking towards him. How much closer would it get before it picked up his scent? He felt the wind was to his face, then he could afford to be patient.

Check Answers

Next snippet ↓

in short/so/then/when

your ideas.

- Words like this can show different relationships between ideas they connect.
These words can show emphasis or sequence of events in time.
How do you think the ideas are connected here?
- Look at the following example; think which of the options expresses the same relationship: The weather was hot. Hence, my ice-cream melted

can you explain

"So" in the sentence functions to show the result or consequence of the conditions described (wind direction and elevation), allowing the character to be patient.

is this cause and effect function?

Yes, in this context, "so" is used to indicate a cause and effect relationship.

Ask a question from the chatbot ...

Send



Broader construct (inferring word meanings) example item

Electric Car

Ford's New All-Electric F-150 Lightning Is More Than It Seems

While driving around suburban Houston during Texas' power crisis in February 2021, which left millions without power for days, Ford energy services lead Ryan O'Gorman was thinking less about that disaster than the next one. Ford at the time was putting the finishing touches on an all-electric version of its best-selling F-150 pickup, with a nifty feature for just such an emergency: in a power outage, the truck would be able to

parked in driveways while I was driving around Texas during that week," says O'Gorman. The F-150 Lightning is the latest offering in Ford's \$22 billion electric vehicle (EV) push. It's also a huge deal for the company which is fighting both fellow incumbents (like GM) and highly-valued upstarts (like Tesla and Rivian) for what analysts

likely to be the future of the auto industry. "We just

What is mentioned as a difference between GM and Tesla?

GM has a more established position

GM is thought to focus on petrol engines

GM keeps fighting fellow incumbents

GM will decide the future of the industry

GM's reputation is more disputed



Mediation in computerised DD-LANG - level 1 mediation

Qu

For

The

Whi

pov

pov

O'G

nex

tou

150

emergency: in a power outage, the truck would be a function as a backup battery, powering an ov house for days. "I can't tell you how many F-150s parked in driveways while I was driving around during that week," says O'Gorman. The F-150 Lightn the latest offering in Ford's \$22 billion electric v (EV) push. It's also a huge deal for the company w fighting both fellow incumbents (like GM) and h valued upstarts (like Tesla and Rivian) for what an say will at least at first be a limited electric market likely to be the future of the auto industry. "W absolutely think there's going to be a big shake

This was not the best option. We will now offer some help and hints...

Sometimes you do not need to know the meaning of an unknown word. However, to know what the word means, it is useful to look at the text around the word (sentence(s), paragraph). If you have a general understanding of what you have read, this will help you figure out what the word might mean.



Show feedback

Sometimes you do not need to know the exact meaning of an unknown word. However, if you need to know what the word means, it is useful to look at the text around the word (sentence(s), paragraph). If you have a general understanding of what you have read, this will help you figure out what the unknown word might mean.



Mediation in computerised DD-LANG - level 2 mediation

The following may help you find the meaning of an unknown word:

- Which word comes right before / after the unknown word? Does that word help you understand the difficult word?
- Can you divide the unknown word into parts? Can you recognise the meaning of the parts? (e.g., un-help-ful)
- Recognising only part of the meaning may help, too. Is the word positive or negative?
- How are details in a sentence or in several sentences connected? For example, the word 'but' suggests that details are contrasted, and 'because' tells us that a reason is given. Look for these kinds of words in the text.

The screenshot shows a user interface for a computerised DD-LANG level 2 mediation task. At the top right, there is a button labeled "Show feedback". Below this, the text "Tesla?" is displayed. A list of five sentences is shown, each in a colored bar. The first two bars are blue, the next two are red, and the last one is blue. The sentences are:

- GM has a more established position
- GM is thought to focus on petrol engines
- GM keeps fighting fellow incumbents
- GM will decide the future of the industry
- GM's reputation is more disputed

A vertical scrollbar is visible on the right side of the list.



Mediation in computerised DD-LANG - level 3 mediation

(note: this level of mediation is item specific, not general / construct specific)

To find the answer to this question, you should think how GM and Tesla are compared to Ford:

- Think what kind of company Ford is.
- Words like "fellow" tell us about GM in relation to Ford.
- "both... and..." signals that GM is a company like Ford whereas Tesla is different.

Remember that the goal is not to understand the full meaning of all the words but only as much as needed to answer the question.

Show feedback

What is mentioned as a difference between GM and Tesla?

GM has a more established position

GM is thought to focus on petrol engines

GM keeps fighting fellow incumbents

GM will decide the future of the industry

GM's reputation is more disputed



Piloting: broad constructs

(1) Human-mediated dynamic assessment

- (a) face-to-face, screen recording, using a script, but deviations allowed
- (b) Specific details, deducing vocabulary, main idea of the paragraph
- (c) 10 upper secondary school students

(2) Scripted mediation online

- (a) screen recorded
- (b) specific details and deducing vocabulary
- (c) asking students about their immediate experiences of the mediation and thinking processes and if they changed during the activity both in writing and more detailed elicitation
- (d) **2 students** who just graduated from upper secondary school

(3) Scripted mediation online

- (a) Written report on experiences with the activity
- (b) **26 MA students** studying different languages

Vastaa pariin lyhyeen kysymykseen: OK

Muuttiko edellisissä kohdissa annettu palaute tapaasi lukea tekstejä? Millä tavoin?

< Keskityin lukemaan vastausvaihtoehtoja tarkemmin ja pyrin etsimään niistä I "pääpointit" >

...



Piloting Revita tasks + mediation with MA students in an assessment course

Challenges

S4: it can feel frustrating to always get the same tips to help them in challenging texts. Perhaps the clues could somehow develop/evolve as the test progresses if the system notices that the same person repeatedly responds incorrectly despite receiving help.

S13: the feedback was always the same, whichever route you chose. This also got a bit boring and the feedback system, which seemed great at the beginning, stopped working. The feedback could be better targeted to the specific question [...].

S9: towards the end, I probably didn't have the energy to concentrate on reading the texts deeply enough to properly weigh the answer options. At this point, it wasn't that I didn't have the skills and strategies to find the necessary information in the texts [...]

Potential solutions / how can AI help with addressing these?



Piloting Revita tasks + mediation with MA students in an assessment course

Challenges

What we thought about

S4: it can feel frustrating to always get the same tips to help them in challenging texts. Perhaps the clues could somehow develop/evolve as the test progresses if the system notices that the same person repeatedly responds incorrectly despite receiving help.

- reformulating / varying mediation
- Switching the focus to the meaning of the strategies and students' recollection of the previous guidance
- mechanism that decides which mediation level is presented for

S13: the feedback was always the same, whichever route you chose. This also got a bit boring and the feedback system, which seemed great at the beginning, stopped working. The feedback could be better targeted to the specific question [...].

- base the mediation on the features of the text, e.g., title, derived words, discourse markers, etc.

S9: towards the end, I probably didn't have the energy to concentrate on reading the texts deeply enough to properly weigh the answer options. At this point, it wasn't that I didn't have the skills and strategies to find the necessary information in the texts [...]

- Instructing learners to take a break when feeling fatigue
- Asking AI to analyse responding behaviour and suggest a break?



Kiitos!
Thank you!

Ari ari.huhta@jyu.fi

Dmitri dmitri.leontjev@jyu.fi

Matt mep158@psu.edu

Roman roman.yangarber@helsinki.fi

WWW: http://r.jyu.fi/DDLANG_en