How time limits shape response processes: Exploring cognitive validity of C-Tests

Anastasia Drackert

Anna Timukova

21st EALTA Conference

30.05.25, Salzburg

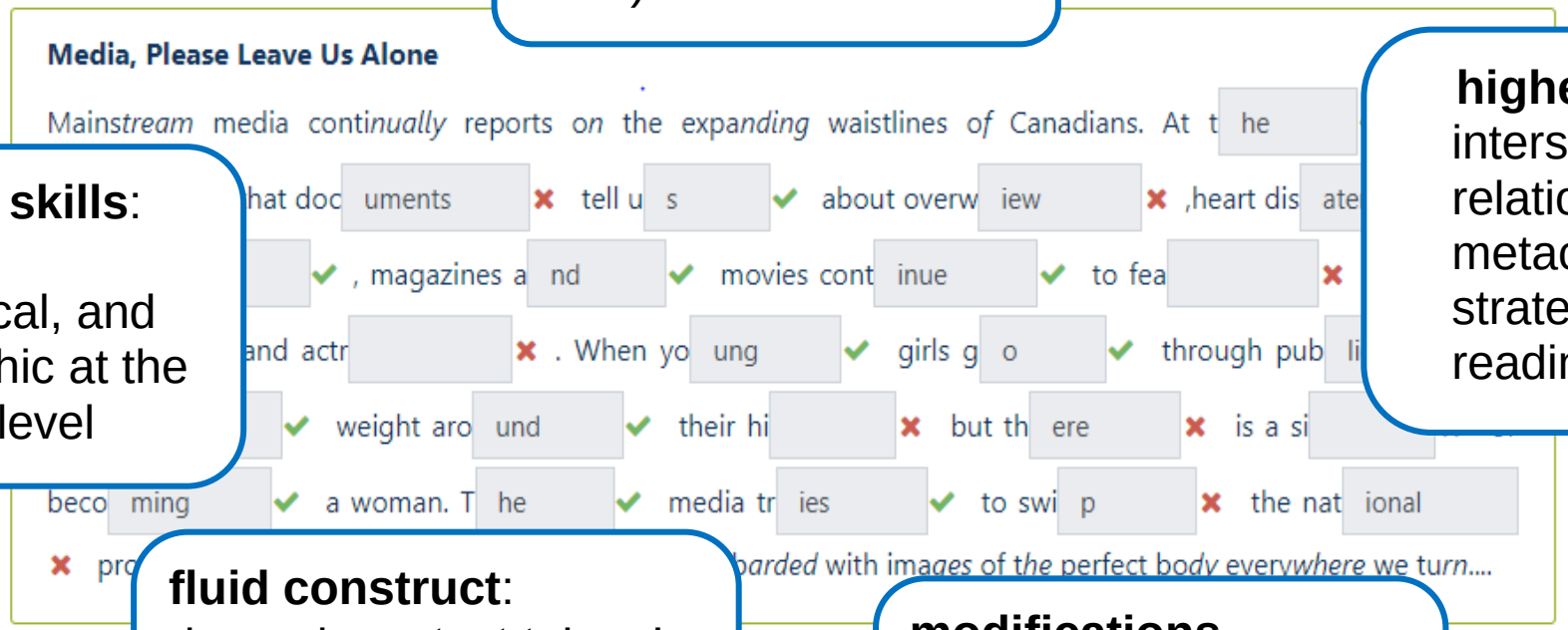
Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

2025 Jahre g.a.s.t.

project number 462766474

C-Test & its construct

general language proficiency (Grotjahn 2012)



low-level skills: lexical, grammatical, and orthographic at the sentence level

higher order skills: intersentential relationships, metacognitive strategies, global reading skills etc.

fluid construct: depends on test takers' proficiency and text characteristics (Sigott 2002; 2006)

modifications possible to construction principles, scoring and **time**

Construct of the speeded C-Test

Canonical C-Test:

5 min per text

amount of **declarative** & **procedural** language **knowledge**

Speeded C-Test:

1:30 – 2:30 min per text

+ **degree of automaticity** of skills and **efficiency** of information **processing**

higher correlations with learners' **writing** and **reading** skills measured under generous time conditions

higher correlations with measures of **listening comprehension** and **speaking** skills (also under **time pressure**)

(Grotjahn, 2010)

Research on the speeded C-Test

Zimmermann (2019): Higher correlations between the speeded C-Test and listening comprehension and speaking (especially dialogic tasks) for B2 learners of GER

Timukova, Möller & Drackert (submitted):
Speeded C-Test scores significantly lower than canonical C-Test scores (by 3.7 points in ENG and 5.1 points (out of 100) in GER);
Speeded C-Test scores predict listening/speaking performance slightly better than canonical C-Test scores
(A2-C1 learners ENG and GER)

Cognitive validity: Looking for (more) *automaticity* in response processes involved in solving *speeded* C-Tests

Participants, instruments, data collection

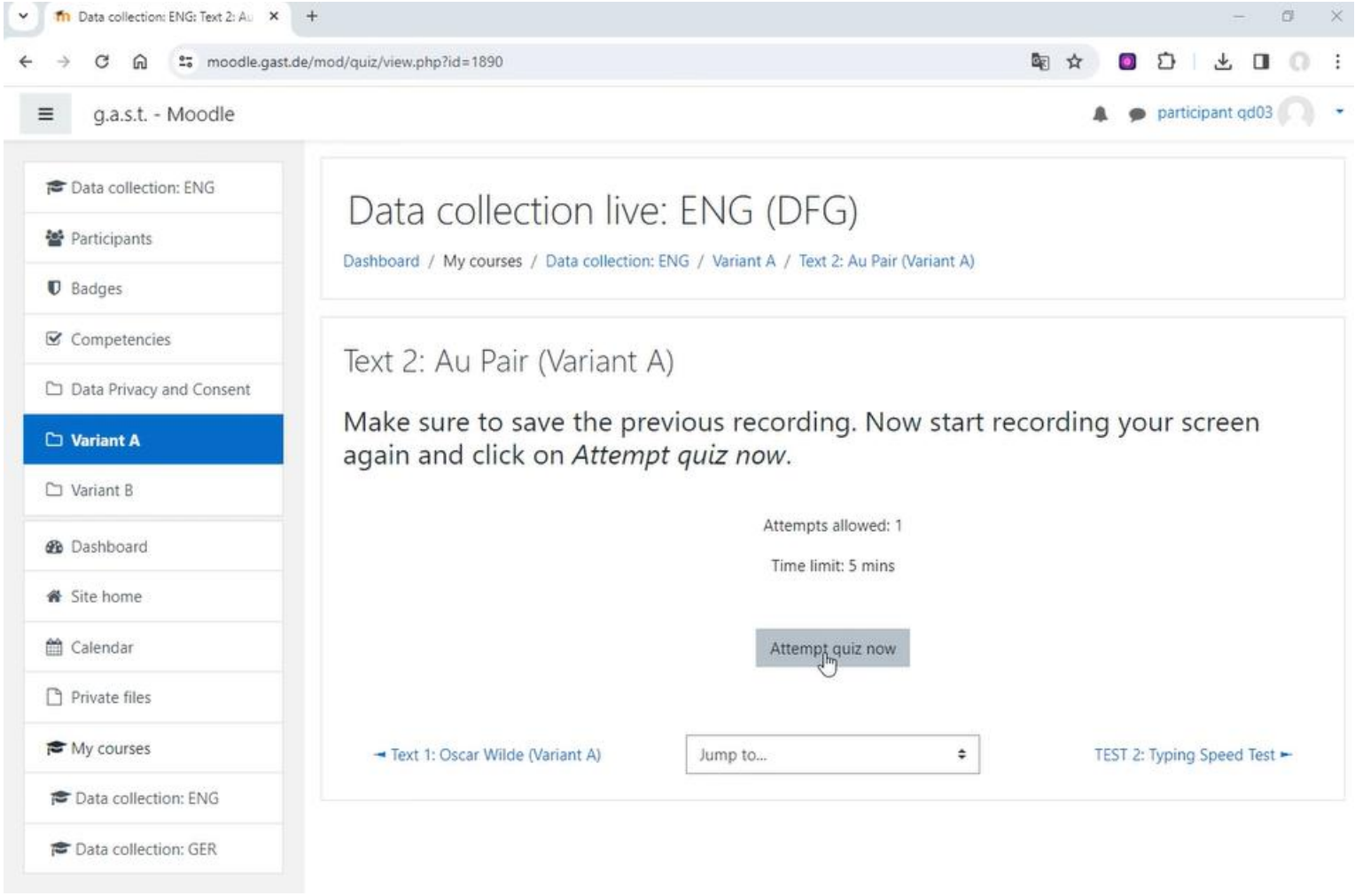
- computer **screen recorded** during test taking *(Wondershare DemoAir*)
- N=**16** GER & N=**25** ENG; level **B2-C1** (ENG more proficient)
- 4 texts* with 20 gaps each in each test version; coded and analysed recordings for **Text 2** (easy) and **Text 4** (difficult)
- time limits: canonical - **5 min** each text; speeded - **1:40** min (Text 2), **2:00** min (Text 4)
- prior text **analysis** – minimal context required: **micro** gaps (gap plus up to two words before and/or after) & **macro** gaps (broader context)

Text 2: Orcas

Reaching up to ten metres in length, the orca is the largest member of the dolphin family. Orcas alw___ live i___ family gro___; they a___ highly soc___ animals. T___ size o___ a gr___ can va___ from ju___ a sm___ number t___ as ma___ as fifty.

**Texts from onSET item bank; comparable difficulty in logits*

Sample recording

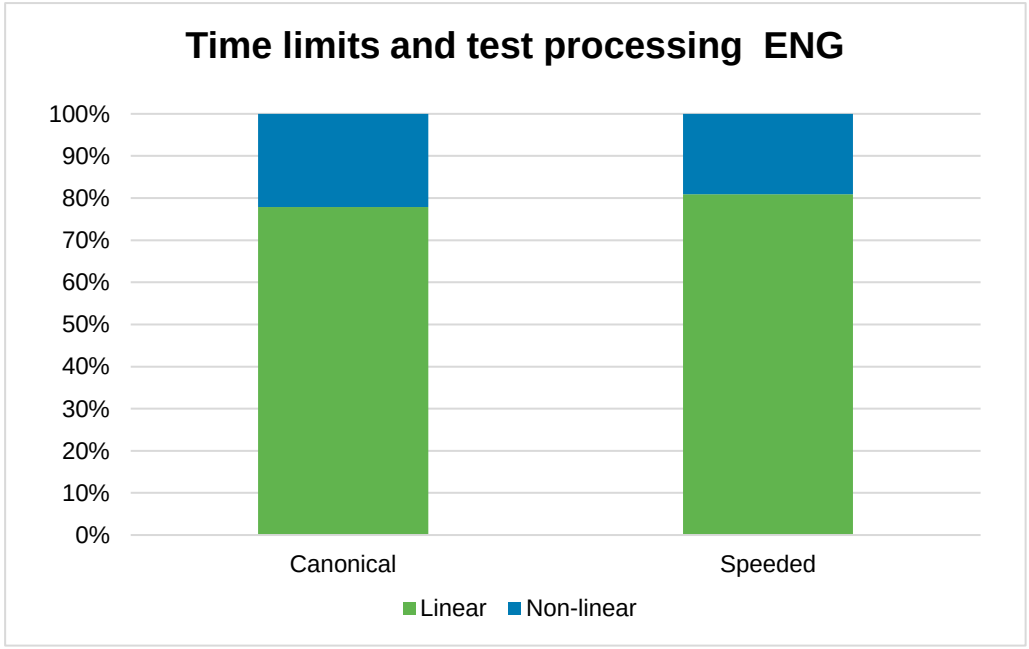


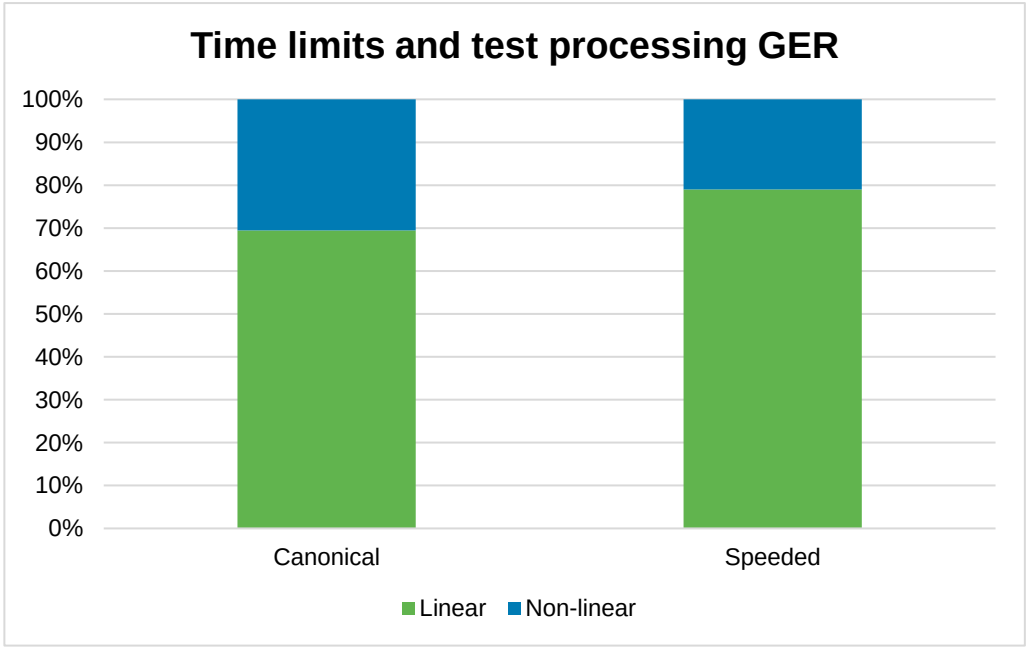
Research questions and findings

2025 Jahre g.a.s.t.

Which time condition elicits more linear processing?*

- **linear processing** dominates in **both** time **conditions** and in both **languages**
- **more** linear processing in **ENG** than in GER
- **more** linear processing in **speeded** than in canonical C-Tests in both languages (more so in GER)

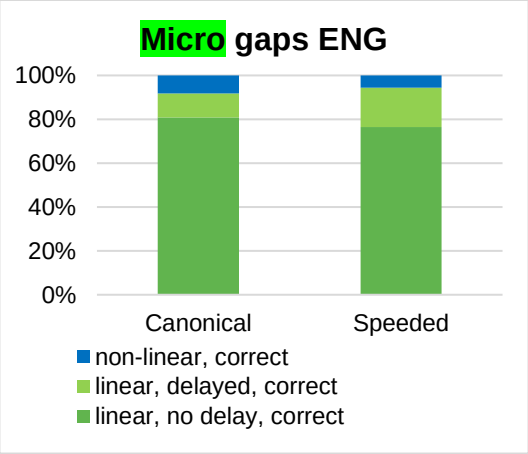


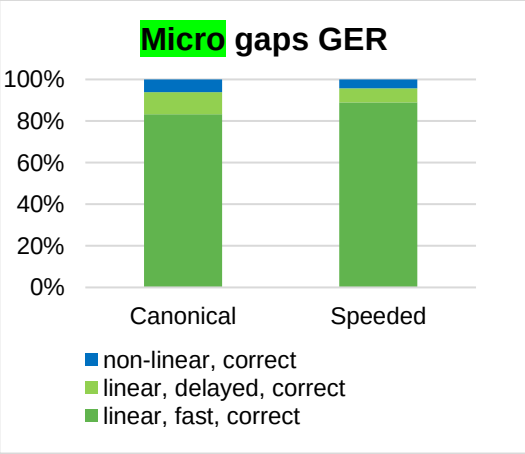


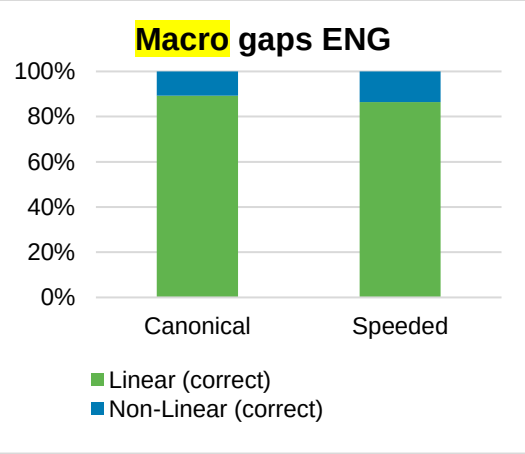
Counted: **all instances of processing (correct and incorrect responses)*

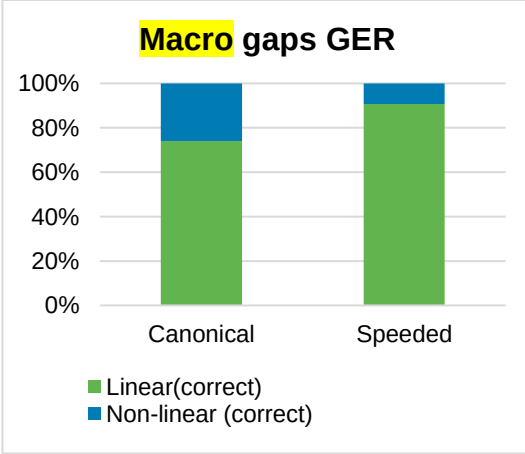
Do gap properties influence response processes?

- **linear** processing dominates across languages, time conditions and **gap types; more** in **ENG**
- slightly **more non-linear** processing of **macro** gaps compared to micro gaps across languages
- **minimal differences** between **time conditions** (except for **macro** gaps in **GER**)







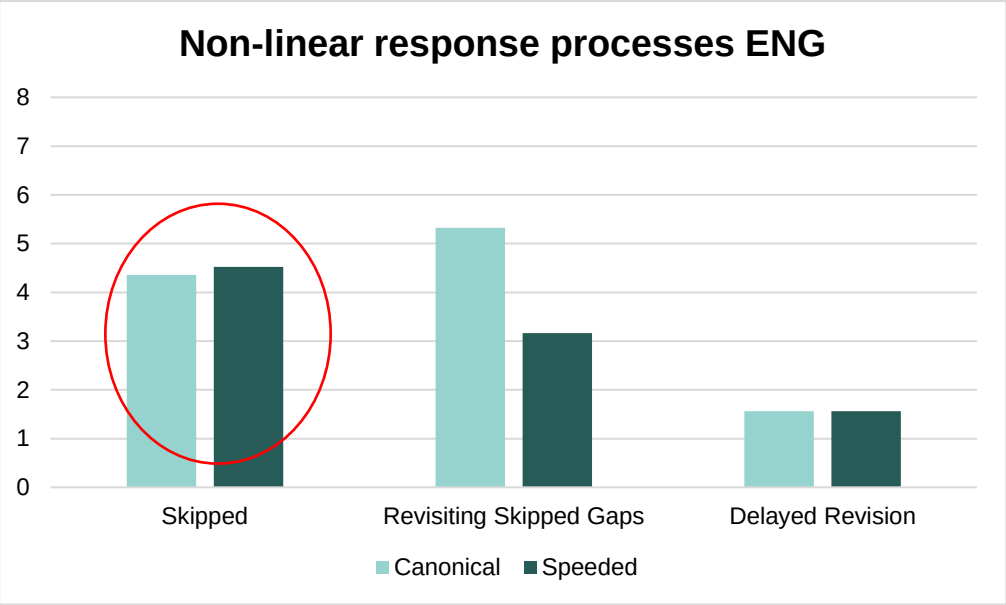


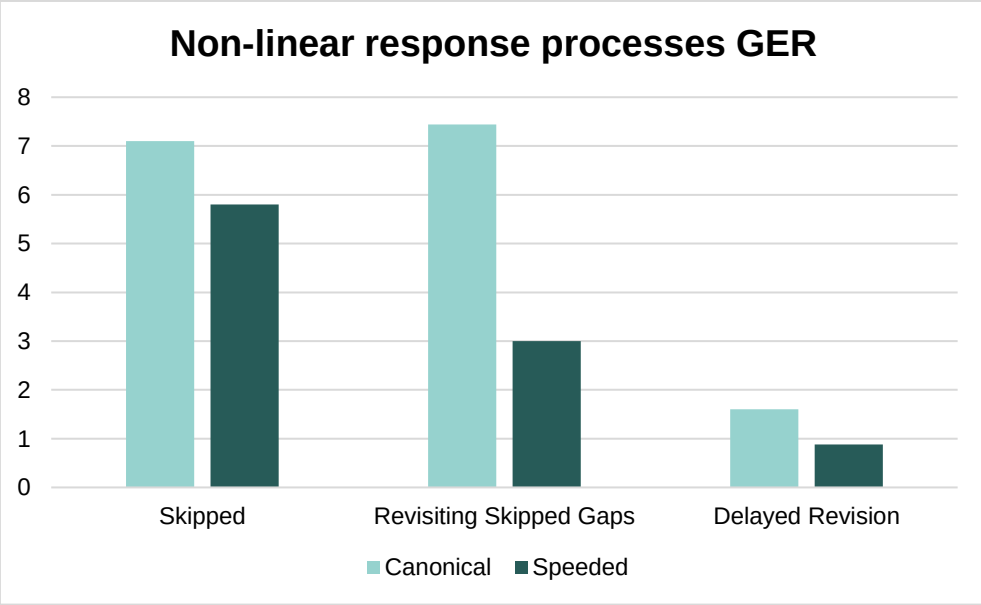
Counted: **instances of processing resulting in correct **responses***

How do non-linear responses differ across time conditions?

Non-linear processing around 20 - 30% (more in canonical C-Tests; most - in GER)

- gaps **skipped** and **revisited more** often in **GER**
- skipped gaps **revisited** more often in **canonical** C-Tests in both languages
- **similar** frequency of **skipping** in speeded and canonical in **ENG**
- **delayed revision** comparatively **rare** across languages and time conditions

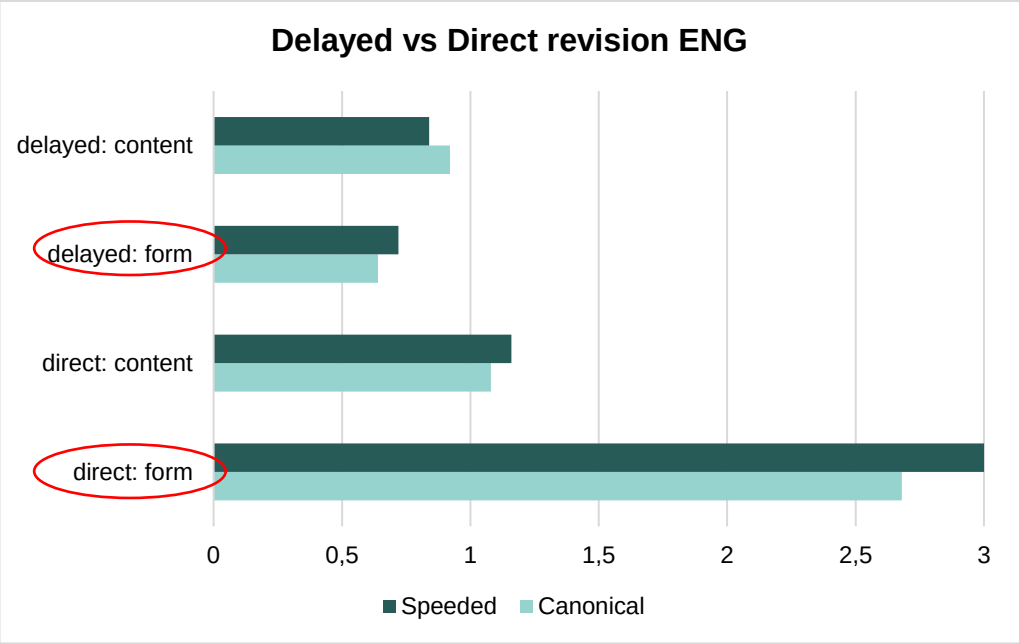


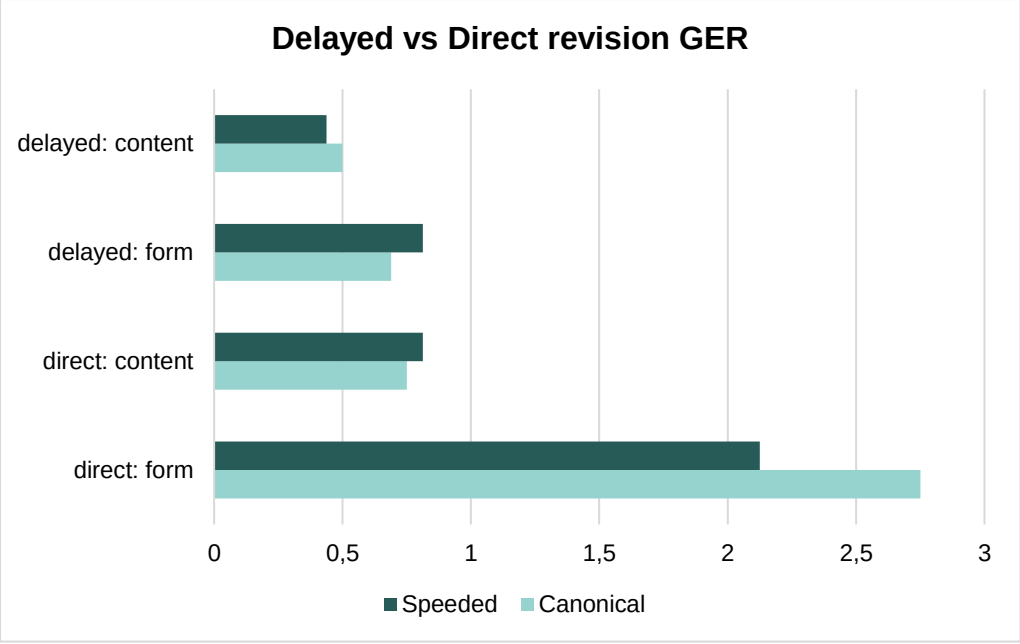


Counted: **all instances of processing (correct and incorrect responses)*

What do test takers correct and when?

- **direct revision** more frequent than delayed revision (both time conditions and languages)
- **revision of form** is more frequent than revision of content (both time conditions and languages)
- **more** direct **revisions** of form in **ENG** than in GER
- **no** clear-cut **differences** between canonical and speeded



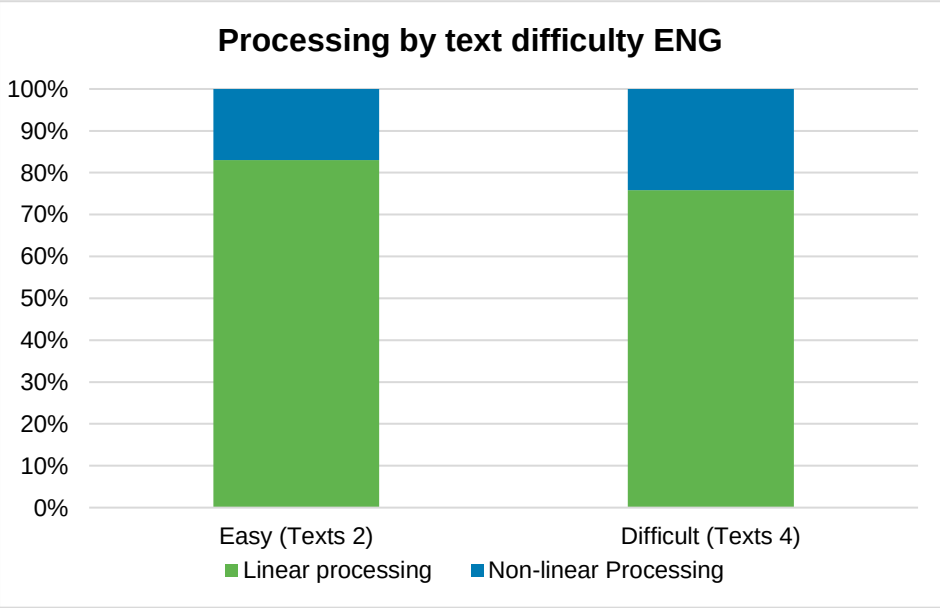


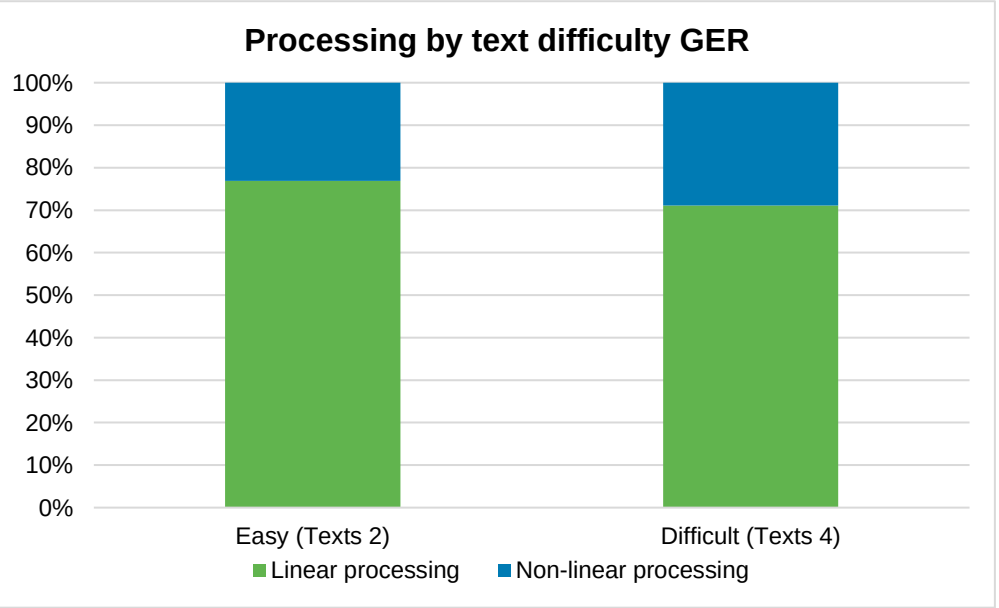
*Counted: ***all instances*** *of processing (correct and incorrect responses)*

How do response processes vary by text difficulty?

- **linear** processing dominates in **both text types** across time conditions and languages
- **more non-linear** processing in **more difficult** texts in both languages (most - Text 4 GER: around 30%)

NB: **more macro** gaps in **Texts 4** than in Texts 2 (ENG: 16 vs 11; GER: 16 vs 12)





Counted: **all instances of processing (correct and incorrect responses)*

Summary and discussion

What is clear:

- **linear** processing **dominates** across time limits, gap properties and texts of varying difficulty
- **more** often **direct** than delayed **revision**; **form** revised **more often** than content

  → C-Tests processed as texts (not language quizzes or puzzles); elicit procedural knowledge
- more **linear** processing in **speeded** than canonical C-Tests
- **difficult texts** and **gaps** requiring **broader context** elicit more **non-linear** response processes

  → reduced time encourages automatic processing

  → more time, gaps requiring broader context and more difficult texts elicit more deliberate, conscious processing (use of declarative knowledge?)

Summary and discussion

Language specific observations, open questions:

- more backtracking in GER
- similar frequency of skipping in speeded and canonical in ENG
- temporal data (time in the gap) – not included; picture not clear (e.g.: more frequent direct revisions – sign of automaticity?)

Important to consider:

- proficiency and proficiency-related strategies
- language differences (syntactical features; 5 “lexical” gaps in ENG; 10 – in GER)

Summary and discussion

Limitations and further research:

- more & more **precise** observations (including accurate temporal data) of response behaviour (e.g. eye-tracking)
- **introspective methods** (e.g. stimulated recall) to explain the behaviour (whether observed through screen recordings or eye-tracking)
- systematic **response analyses** to take into account linguistic features
- **proficiency** included as a variable (but also personality traits)

Thank you!