

Descargar estas diapositivas ↻

<http://bit.do/asefie1>



# Desarrollo de la fluidez oral en lengua extranjera: experimento de medición semiautomática de los efectos de aprendizaje

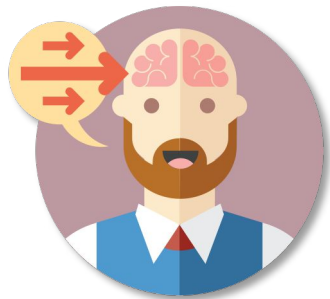
Serge Bibauw<sup>1,2,3</sup> Louis Escouflaire<sup>3</sup> Thomas François<sup>3</sup> Piet Desmet<sup>2</sup>

<sup>1</sup> Universidad Central del Ecuador

<sup>2</sup> ITEC, imec research group at KU Leuven

<sup>3</sup> CENTAL, UCLouvain

# Fluency development in a foreign language: semi-automated measurement of learning gains



## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

Utterance fluency, fluency metrics and evaluation

## Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

## Results and discussion

Fluency metrics correlated with proficiency  
Short-term treatment effect on fluency  
Limits and perspectives

# Fluency development in a foreign language: semi-automated measurement of learning gains

## ► Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

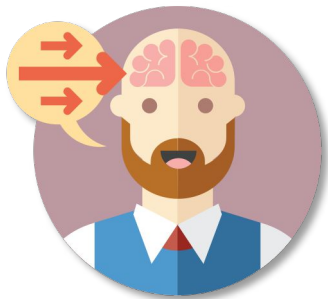
Utterance fluency, fluency metrics and evaluation

## Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

## Results and discussion

Fluency metrics correlated with proficiency  
Short-term treatment effect on fluency  
Limits and perspectives



# Theoretical context

## Second language acquisition and testing

*Knowledge-based approach*

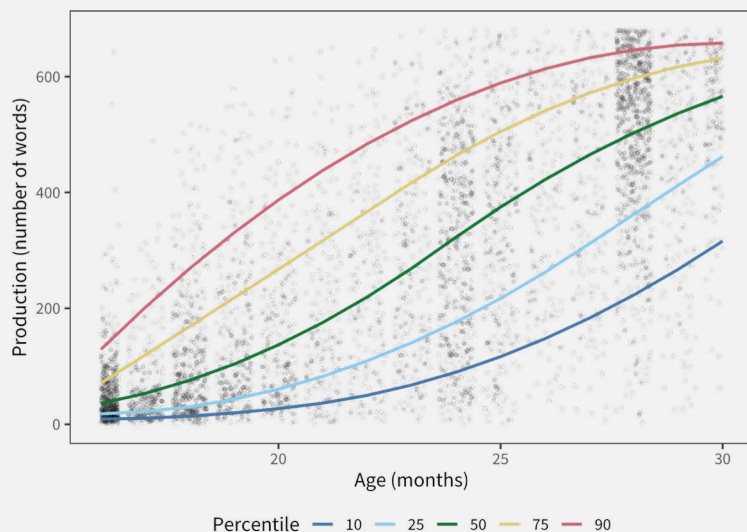
**Knowledge (Declarative)**

→ **Tests**

⇒ **Vocabulary size test**

(very efficient proxy of proficiency)

(Milton, 2013)



*Task-based approach*

**Skill (Proficiency / Procedural)**

→ **Performance**

(ability to express something)

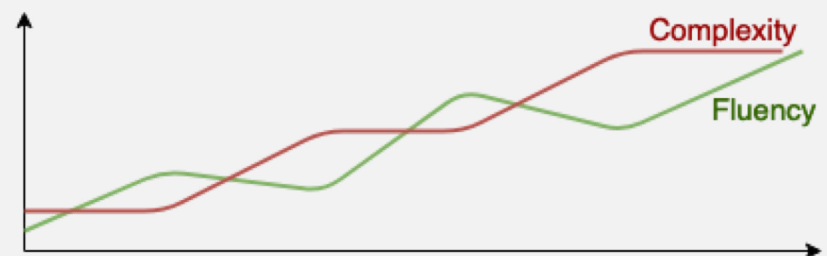
Proficiency

as

**Complexity + Accuracy + Fluency**

(Housen et al, 2012; Michel, 2017)

Trade-off between complexity, accuracy & fluency (theory)



# Research context

## Effects of dialogue systems on fluency dev.

Interactive practice with a **dialogue system**, in this case, integrated within an educational game.



Objective: develop **fluency** in A1-A2 learners.

Randomized controlled experiment with young Flemish learners of French (N=215) in Belgium

# Research context

## Dialogue systems for language learning

(Bibauw, François & Desmet, 2019)

Any application or system allowing

to maintain a **dialogue**

[ immediate, synchronous interaction ]

[ written or spoken ]

with an **automated agent**

[ chatbot, talking robot, automated personal assistant, conversational agent, non-player character in a video game... ]

[ tutorial CALL (≠ computer-mediated communication) ]

for **language learning** purposes.

Logged in as sbibauw

Logout

Target language:

fr

Tutor language:

en

Interface language:

Réglages

# Language Hero

## Conversations:

Conversation 1: After the storm - Meet Sensei and find out what happened and where you are.

Meilleur score: 828

Conversation 2: Meet Baldog - Meet Baldog and ask him for help.

Meilleur score: 0

Conversation 3: The snails - Vincent - Get to know the snails family

Meilleur score: 426

Conversation 4: The snails - Angélique - Get to know the mother of the snails family

Meilleur score: 0

Conversation 5: The snails - Claudette - Get to know one of the triplets of the snails family

Meilleur score: 0

Conversation 6: Return to Baldog - Go back to Baldog and tell him his problem is solved.

Visit the world

Quit

Conversation: The snails - Vincent - Get to know the snails family



Contextualization

- ▶ *He: Bien le bonjour! Comment t'appelles-tu?*
- ✔ *You: bonjour je m'appelle Marco*
- ▶ *He: Enchanté de faire ta connaissance, Rinc! Rinc. Rinc. Rinc. Ne t'en fais pas, je ne suis pas fou. C'est juste que je répète ton nom pour ne pas l'oublier.*
- You: Comment tu t'appelle?*
- He does not seem to have heard you...*
- You: Tu t'appelle coment?*
- He does not seem to have heard you...*
- ✔ *You: Tu t'appelle comment?*
- Correction: appelle - Vérifiez l'accord entre le pronom « Tu » et le verbe « appelle ».*
- Task accomplished: Good. That was what we were wondering about.*
- ▶ *He: Moi, c'est Vincent. Elle, là-bas, c'est Angélique. Ça, c'est Delphine. Puis on a Georges dans le coin. Et évidemment, on ne peut pas oublier les triplées : Lisette, Claudette et Yvette. Oh! Et puis le petit là-bas, c'est Louis.*

Corrective feedback

Score: 405 ?  
Friendship lvl0: Acquaintance

Gamification

🔥 Current task (2/30):  
Say it is nice to meet them.

Microtasks to guide the conversation

Type or say your answer:

Type text..

Free written input

➔ Send your reply

🎤 Record your answer

? Disable help

⊗ End conversation

We can give you suggestions you can use to come up with an answer:

Scaffolding



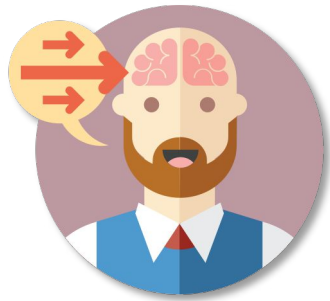
# Fluency development in a foreign language: semi-automated measurement of learning gains

## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## ► Measuring fluency development

Utterance fluency, fluency metrics and evaluation



## Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

## Results and discussion

Fluency metrics correlated with proficiency  
Short-term treatment effect on fluency  
Limits and perspectives

# Theory and state of research

## Fluency (Segalowitz, 2010)

Speaking fluency as a multidimensional construct

- **Cognitive fluency** (skill-level)
  - no direct access
- **Utterance fluency** (performance-level)
- **Perceived fluency** (listener perspective)

# Theory and state of research

## Utterance fluency (Segalowitz, 2010, 2017)

- **Speed fluency**
  - speech rate, articulation rate, syllable duration, length of runs (syllables), duration of runs (sec)... (Bosker et al, 2013; Hilton, 2014; Kormos & Denes, 2004; Götz, 2013...)
- **Breakdown/Pauses**
  - silent pause rate, silent pause duration... (Bosker et al, 2013; de Jong & Bosker, 2013; Kahng, 2014; Hilton, 2014...)
  - ~~filled pauses~~: not good differentiator (Cucchiarini et al, 2002...),  
unrelated to other fluency measures (Segalowitz et al 2017)
- **Repair fluency**: not good differentiator of proficiency (Cucchiarini et al, 2002; Revesz et al 2016; Saito et al 2018; Dumont, 2017...)

# Theory and state of research

## Fluency metrics

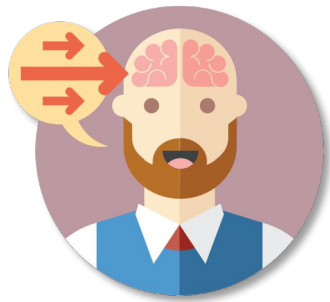
Dozens of possible metrics

Combined with dozens of different operationalizations:

- **silent pause threshold: in general 250ms** (de Jong & Bosker, 2013; Préfontaine et al, 2016)
- pruning and inclusion criteria for syllables and words
- syllables count
- normalization
- combinations of different denominators, order, etc.
- logarithmic transformations

⇒ Need to **compare these operationalizations**, not only theoretically, but in terms of **empirical adequacy** with the metrics' **purpose** (here: measure language development)

# Fluency development in a foreign language: semi-automated measurement of learning gains



## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

Utterance fluency, fluency metrics and evaluation

## ► Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

## Results and discussion

Fluency metrics correlated with proficiency  
Short-term treatment effect on fluency  
Limits and perspectives

# Methods Procedure

1-4 weeks,  
depending  
on school  
schedule

All sessions  
at school

## Pretest

- Computer-delivered spoken interview**
- Target vocabulary test
- Vocabulary size test**

In-app session (max 50 min): **DSys / DCompl**

In-app session (max 50 min) : **DSys / DCompl**

In-app session (max 50 min) : **DSys / DCompl**

## Posttest

- Computer-delivered spoken interview**
- Perceptions questionnaire
- Target vocabulary test

# Methods

## Population and group assignment

4 schools volunteered to participate, with 2-3 classes each:

$$N_{\text{clusters}} = 11$$

$$N_{\text{participants}} = 215 \text{ (208 complete cases)}$$

**Random assignment** of classes to 3 conditions (distr. equally across schools):

- **Dialogue System** (experimental):  $n_{\text{D.Sys.}} = 81$
- **Dialogue Completion** ('baseline'):  $n_{\text{D.Compl}} = 79$
- **Control** ('business-as-usual')  $n_{\text{control}} = 49$

Flemish 2<sup>nd</sup> year secondary school learners of French ( $M_{\text{age}} = 13.4$  y.o.)

L1 = 95,3 % Dutch

L2 = French = first L2,  $M = 3,1$  years of instruction, mostly at **A1** level

( $M_{\text{score}}$  in productive vocabulary size test = 3.6/30 in 1K frequency band)

10 (near-)native speakers of French excluded (final N = 198)

Intervention · Dialogue system

*LanguageHero*, dialogue-based game for young learners

16

Codeveloped with Leuven-based start-up *Linguineo*.

(Main) target audience: teenagers (10-14).

Prototype developed for French for Dutch-speaking learners.

Task-based free conversational written interaction.



# Intervention · Conditions

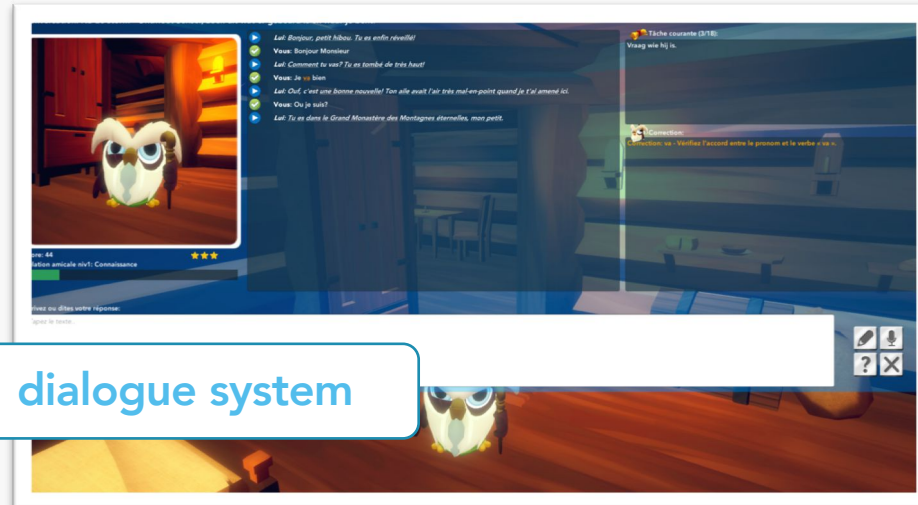
## Interactive vs. static dialogue

Compare:

(A) fully interactive,  
immediate/synchronous  
**dialogue system**

(B) classic, asynchronous  
**dialogue completion task**

Conditions with identical tasks,  
input, output opportunities,  
feedback and scaffolding.

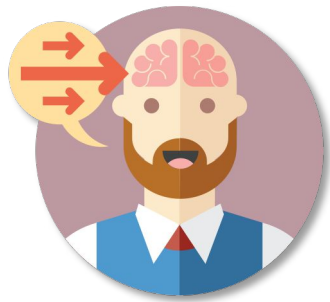


**dialogue system**



**dialogue completion**

# Fluency development in a foreign language: semi-automated measurement of learning gains



## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

Utterance fluency, fluency metrics and evaluation

## ► Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

## Results and discussion

Fluency metrics correlated with proficiency  
Short-term treatment effect on fluency  
Limits and perspectives

# Computer-delivered speaking interview

## Automated speaking test

Individual, in-class &  
simultaneous, with headset

## 24 questions

from basic ("How are you?") to questions targeting  
specific communicative functions ("Can you  
describe your French teacher?")

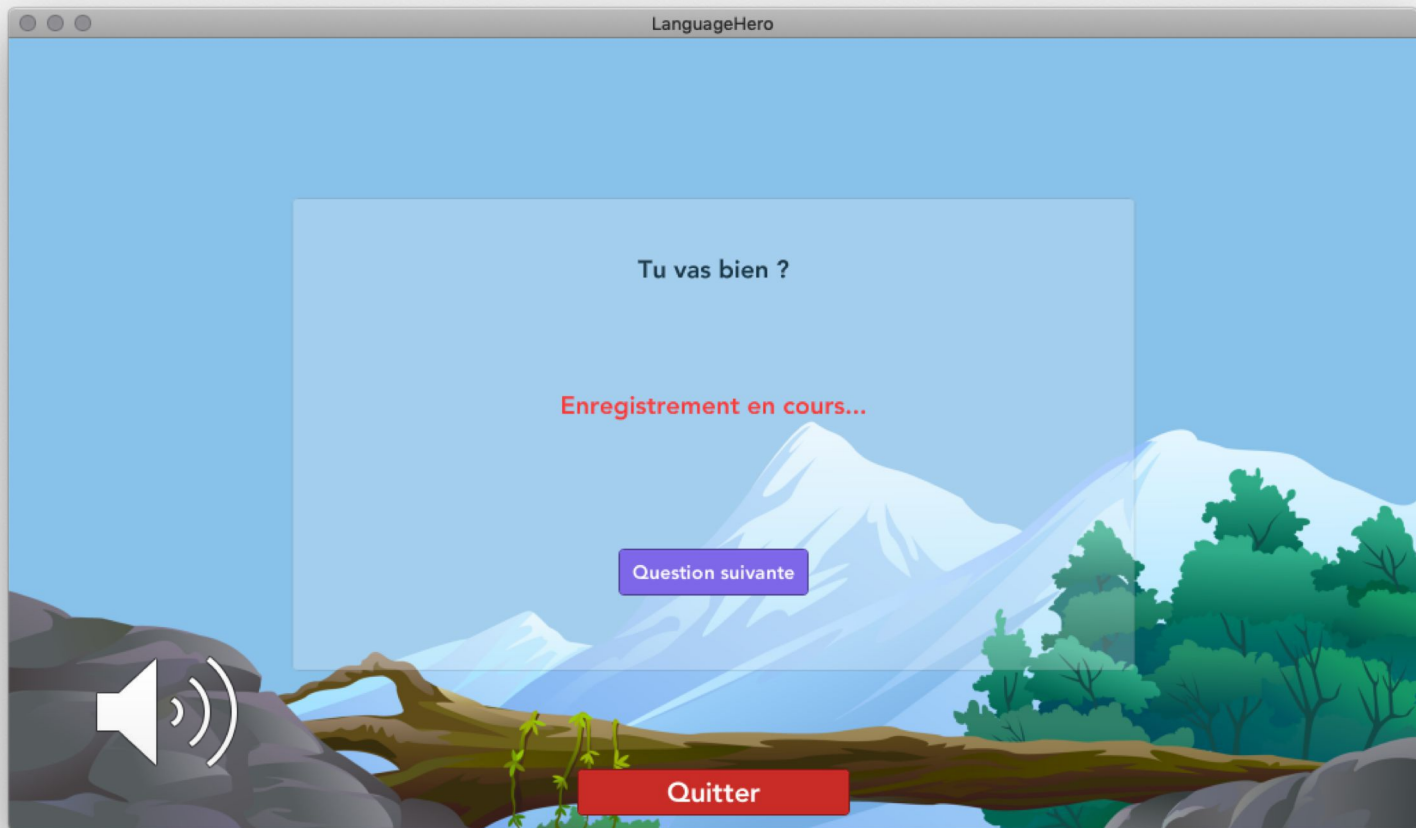
## Question oral + written presentation,

then **automatically starts recording**,  
30 sec limits or "Next question" button



Methods · Instruments

# Computer-delivered speaking interview



## Vocabulary Size Test

### Productive Vocabulary Size Test

Developed and validated for VocabLab project  
(Peters et al, 2019a; Noreillie, 2019)

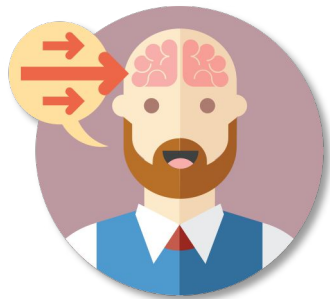
Gap-filling in L2 with given first letter + L1  
translation (Productive)

60 items (< frequency bands 1K + 2K)

Computer-delivered, made adaptive  
(30 1K items, then if > 50% correct: + 30 2K items)

Used as a proxy of L2 proficiency (at pretest only)  
(used as covariate in MEM)

# Fluency development in a foreign language: semi-automated measurement of learning gains



## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

Utterance fluency, fluency metrics and evaluation

## Methods

- ▶ Computer-delivered spoken interview  
Automated analyses for fluency measurement

## Results and discussion

Fluency metrics correlated with proficiency  
Short-term treatment effect on fluency  
Limits and perspectives

# Methods

## Processing of spoken responses

±11 000 single audio files (N=215 \* 24 questions \* pre+post)

- **Automated speech recognition** (Google Cloud Speech-to-text) for transcription
- Manual **correction** of transcriptions +
- **Annotation** of filled pauses, L1/LF use, meta-discourse, etc. with tagging layer
  - allowed to then include/exclude certain features for metrics variants

# Methods

## Computation of fluency metrics

- Automated detection of pauses (Praat syllable nuclei detection script, de Jong & Wempe, 2009)
- Alternate methods for silent pause detection, and syllables/length count.
- Automated computation of syllables from transcript, with variations in pruning.
- Computation of all possible variants of every temporal fluency metric.



## Methods

# Composite fluency index

To obtain a single, aggregate/composite index of temporal utterance fluency:

- **Principal Component Analysis (PCA)**
- Selecting first component  
(76% of variance explained)
- Checking loadings of most important fluency variables

# Fluency development in a foreign language: semi-automated measurement of learning gains



## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

Utterance fluency, fluency metrics and evaluation

## Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

## ► Results and discussion

Fluency metrics correlated with proficiency  
Short-term treatment effect on fluency  
Limits and perspectives

# Results

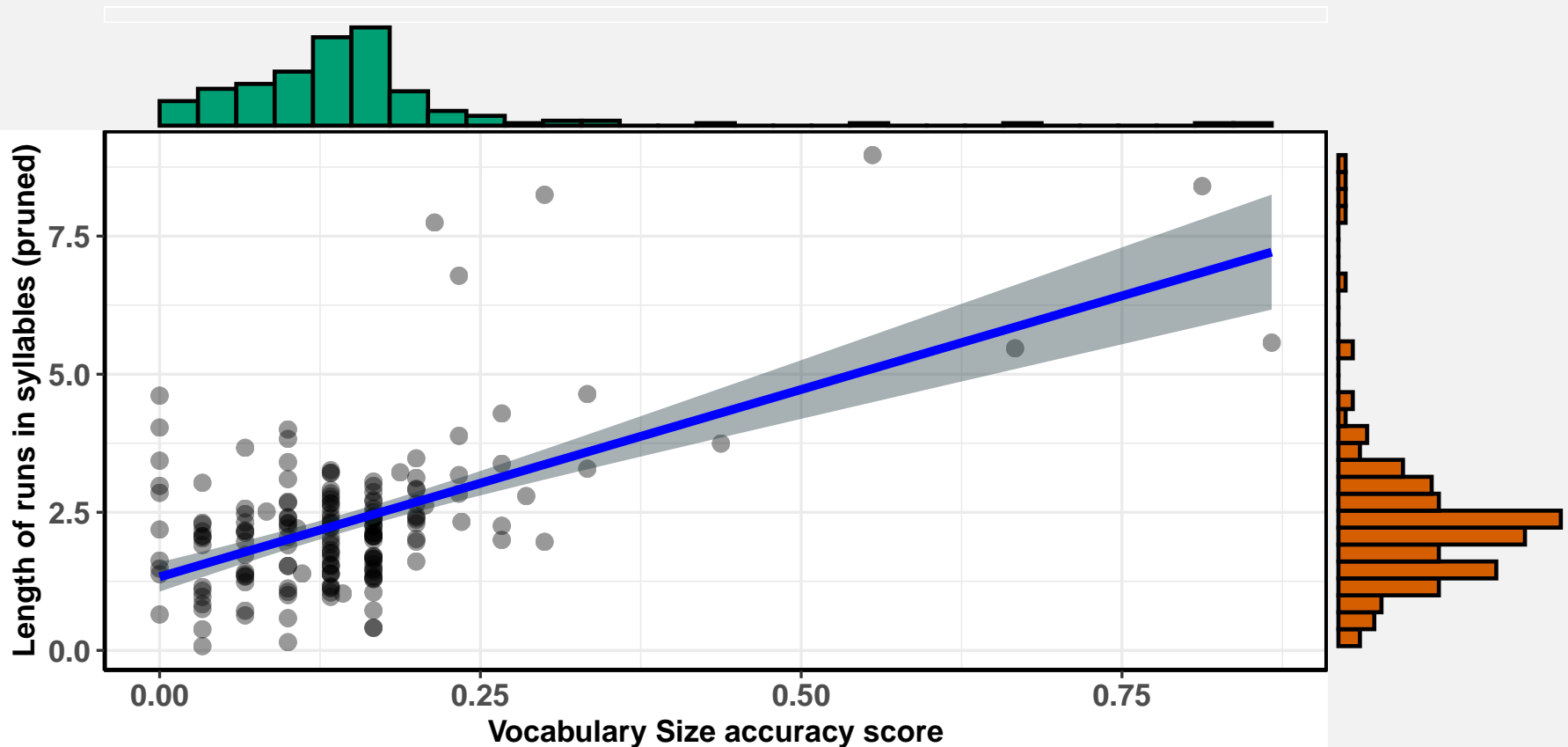
## Fluency metrics

Variable	→ Correlation with Vocabulary Size Test	<i>r</i>
<b>Length of runs in syllables (pruning all proper nouns)</b>		<b>0,58</b>
Length of runs in syllables (pruning non target)		0,57
Length of runs in syllables (no pruning)		0,57
Length of runs in syllables (alternate syllable count)		0,56
<b>Speech rate (pruning all proper nouns)</b>		<b>0,55</b>
Speech rate (no pruning)		0,53
<b>Number of syllables (pruning all PN)</b>		<b>0,46</b>
Number of words (pruning all PN)		0,45
<b>Articulation rate (inverse syllable duration)</b>		<b>0,43</b>
<b>Length of runs in seconds (pruning)</b>		<b>0,36</b>
Speech/Time ratio		0,26

# Results

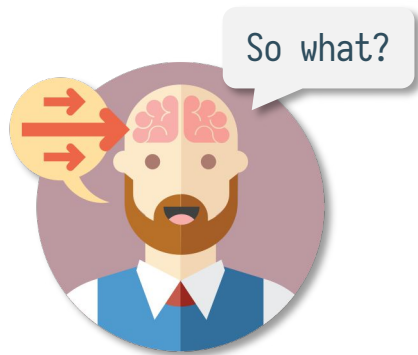
## Fluency metrics

$t(172) = 9.38, p = < 0.001, r_{\text{Pearson}} = 0.58, \text{CI}_{95\%} [0.47, 0.67], n_{\text{pairs}} = 174$



In favor of null:  $\log_e(\text{BF}_{01}) = -32.83, r_{\text{Cauchy}}^{\text{JZS}} = 0.71$

# Fluency development in a foreign language: semi-automated measurement of learning gains



## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

Utterance fluency, fluency metrics and evaluation

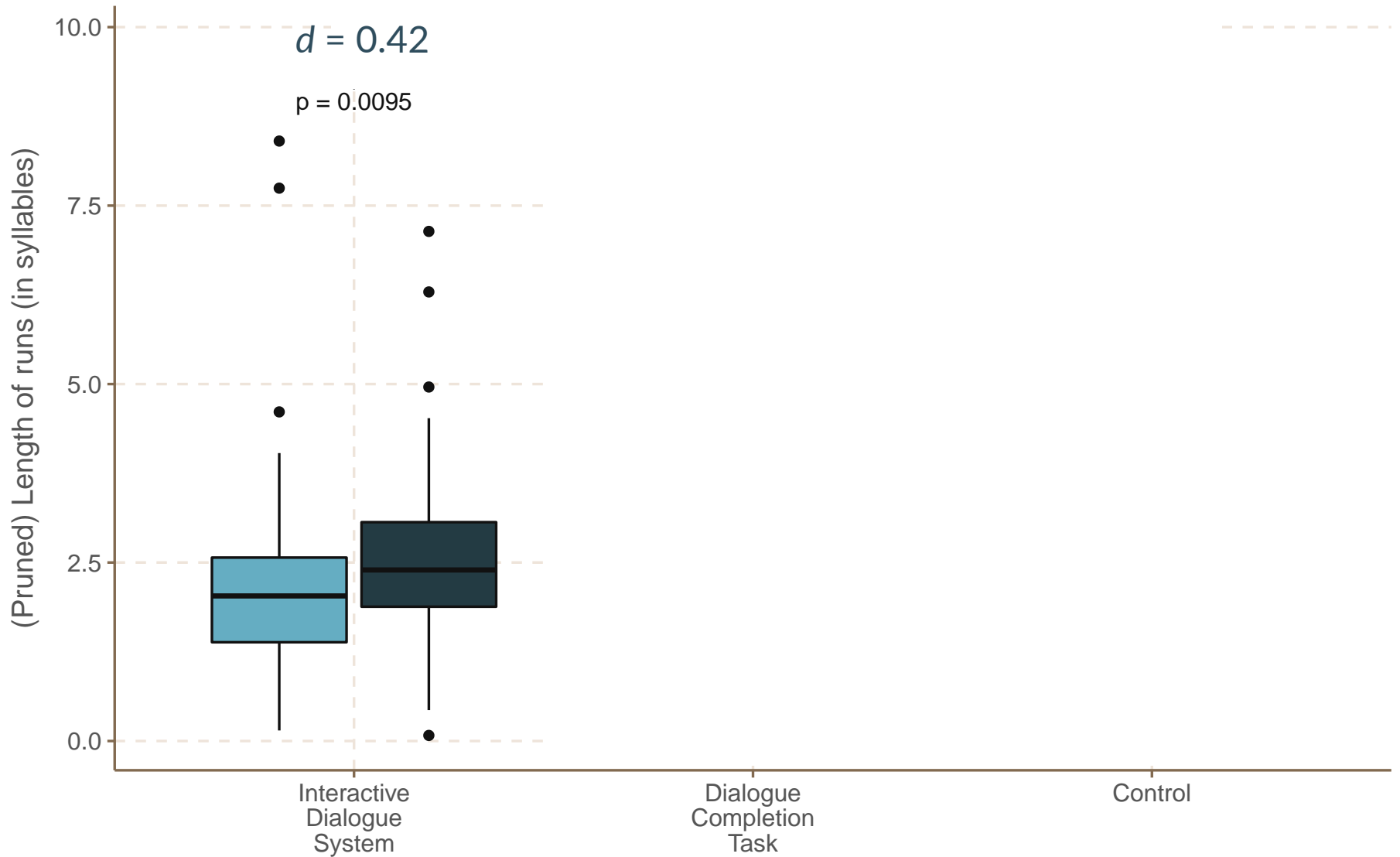
## Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

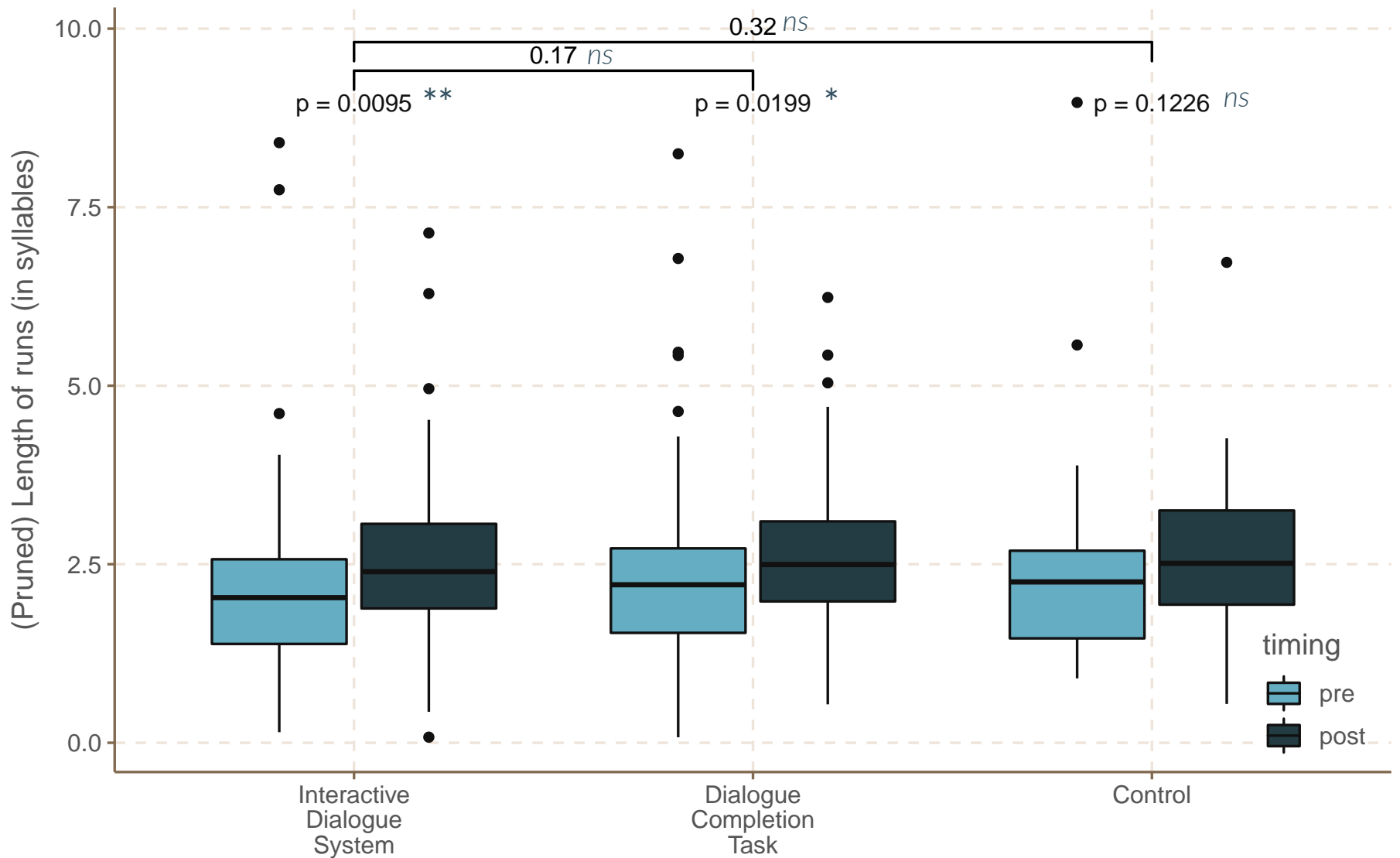
## Results and discussion

- ▶ Fluency metrics correlated with proficiency
- Short-term treatment effect on fluency
- Limits and perspectives

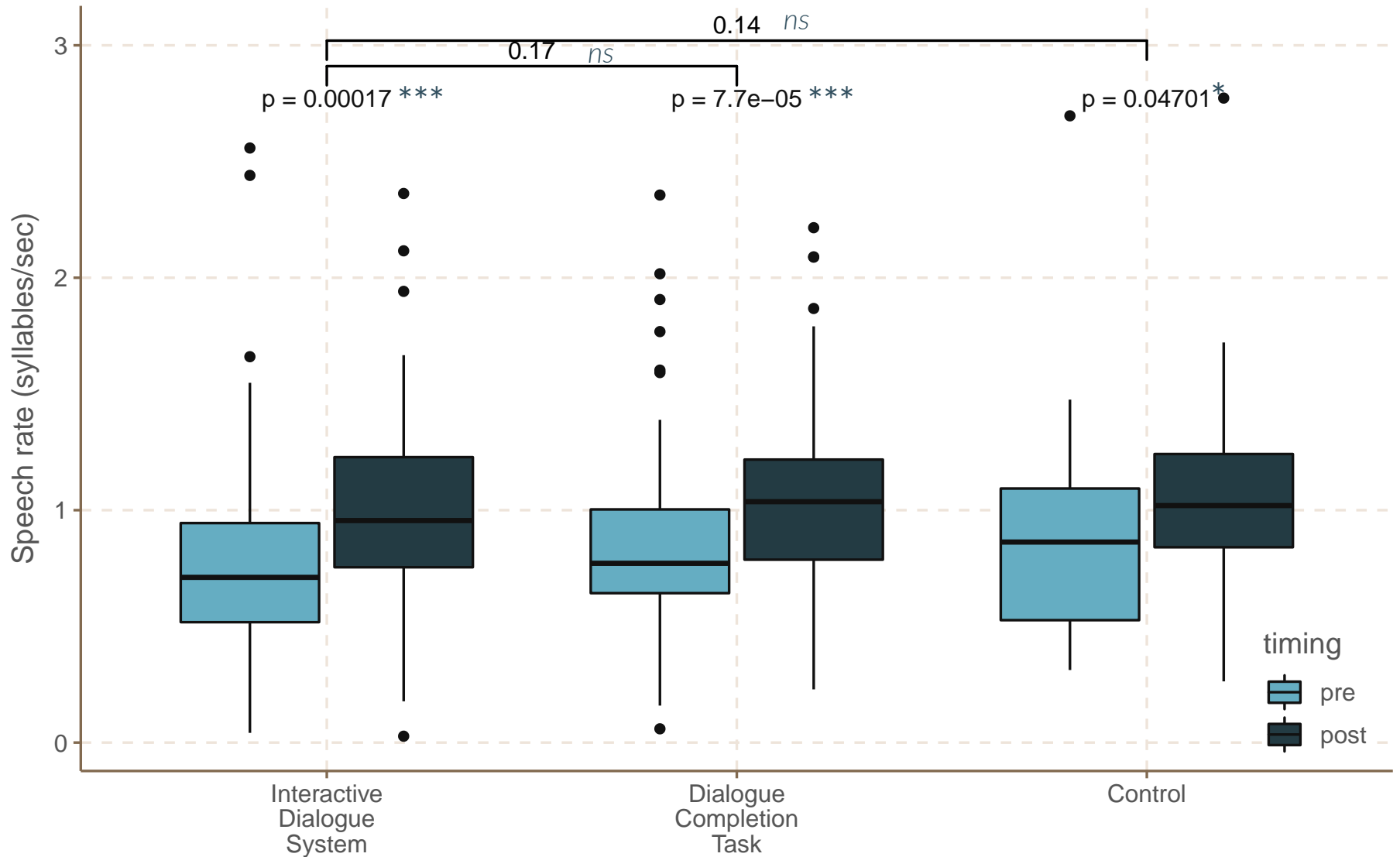
# Results: effects on **Fluency** (length of runs)



# Results: effects on Fluency (length of runs)

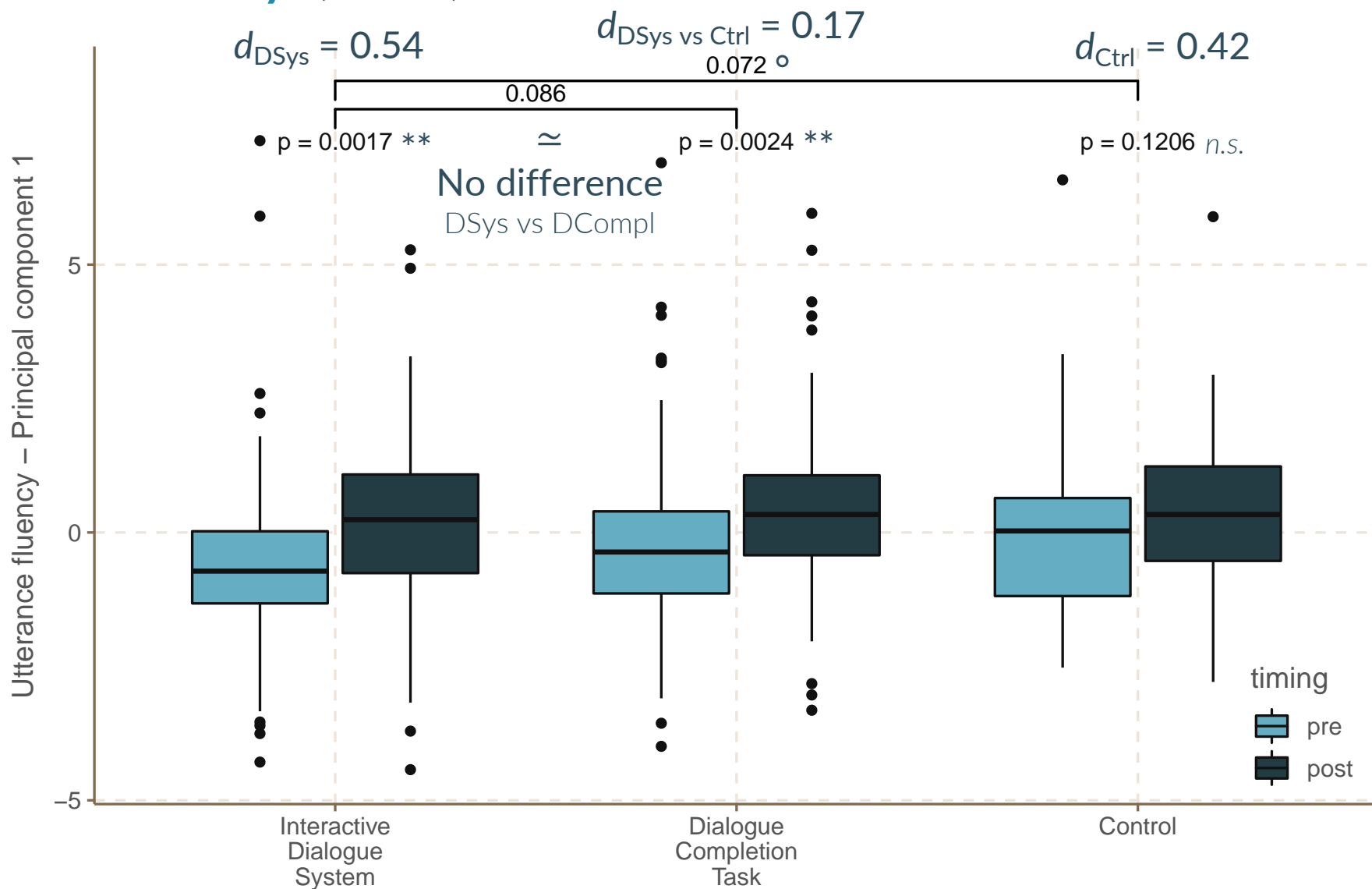


# Results: effects on Fluency (speech rate)





# Results: effects on Fluency (PC1)



# Discussion

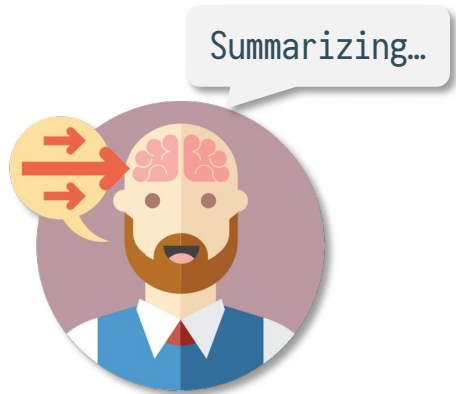
## Fluency

**Very small** effect ( $d_{\text{DSys vs Ctrl}} = 0.17$ ), when controlled for “base development” and training to the test effect,

but very **short treatment** (2h) → expected (effect on general L2 speaking proficiency by *written* practice)

No difference between interactive and non-interactive system.

# Fluency development in a foreign language: semi-automated measurement of learning gains



## Context

Performance-based proficiency assessment  
Dialogue systems for language learning

## Measuring fluency development

Utterance fluency, fluency metrics and evaluation

## Methods

Computer-delivered spoken interview  
Automated analyses for fluency measurement

## Results and discussion

- Fluency metrics correlated with proficiency
- ▶ Short-term treatment effect on fluency
- Limits and perspectives

Conclusions

## Effects of dialogue-based CALL

Very small effect on **fluency**

Still quite promising that possible to observe an effect on fluency on such a small timeframe.

Perspectives

## Automated speaking fluency testing

Fine-grained evaluation of fluency metrics via automated comparison

Simultaneous individual speaking test for >30 learners

Precise automated recording of fluency variables

Almost fully automated processing pipeline

⇒ **Methodological innovation**

# Perspectives

## Dialogue systems as a research environment

Dialogue systems offer **fully controllable and reproducible interaction**: opportunities to monitor and to alter infinity of details.

Experimental testing (A/B testing) with different types of tasks, instructions, feedback, exposure, reactions...

→ Opportunity to compare **writing fluency and speaking fluency** in similar settings

Thank you!  
Merci !  
Dank u!  
¡Gracias!

Serge Bibauw

[sbibauw@uce.edu.ec](mailto:sbibauw@uce.edu.ec)

Louis Escouflaire

Thomas François

Piet Desmet

Descargar estas diapositivas 

<http://bit.do/asefie1>



More info: <https://serge.bibauw.be>