Double play listening assessment: towards increased authenticity

Franz Holzknecht

Lancaster University &
University of Innsbruck

EALTA 2019, Dublin

Overview

- Background
- Study 1 (large-scale)
- Study 2 (verbal recall)
- Findings: response processes
- Implications

International high-stakes listening tests

Double play only

Cambridge, Abitur, Baccalauréat, Matura, …

Double play (or more) optional

Aptis, Oxford Test of English, …

Mixture of single play and double play

Trinity, PTE General, Test DaF, ...

Single play only

IELTS, TOEFL, TOEIC, PTE Academic, GEPT, …

Previous research on double play (1)

Majority of studies…

- only investigated whether double play impacts test scores

(Berne, 1995; Cervantes & Gainer, 1992; Chang & Read, 2006, 2007; Field, 2015; Horness, 2013; Iimura, 2007; Lund, 1991; Otsuka, 2004; Ruhm et al., 2016; Sakai, 2009)

However…

More relevant:

- effects of double play on coverage of the listening construct

Field (2015): increased higher-order processing and lower anxiety in second play vs. the first in a double play condition

Research question

What are the differences in test takers' **response processes** between listening tasks completed in **single play** and **double play**?

Response processes:

construct-relevant	construct-irrelevant
cognitive processes	test-taking strategies
listening strategies	anxiety

Study 1 – Overview

- Four listening tasks
  - Austrian Matura exam
  - Two multiple-choice and two open format tasks (B2)
  - Potential confounding factors controlled for
- Questionnaire
  - Listening strategies
  - Test-taking strategies
  - Anxiety
- Statistical analyses of questionnaire responses

Study 1 – Research design

# of times heard	version 1	version 16
1	MC 1	OF 2
1	MC 2	OF 1
	Quest.	Quest.
2	OF 1	MC 2
2	OF 2	MC 1
	Quest.	Quest.

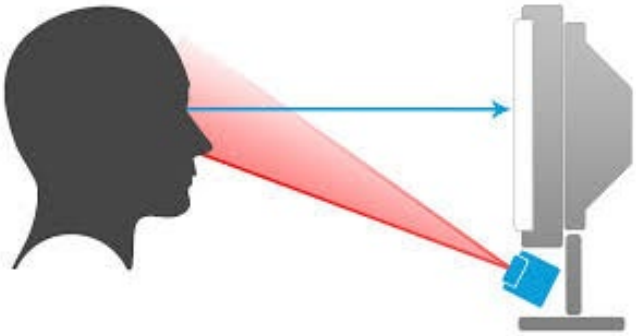
→ 16 different versions

- 16 school classes (one version per class)
- 16-18 year-old Austrian school students
- 306 participants overall

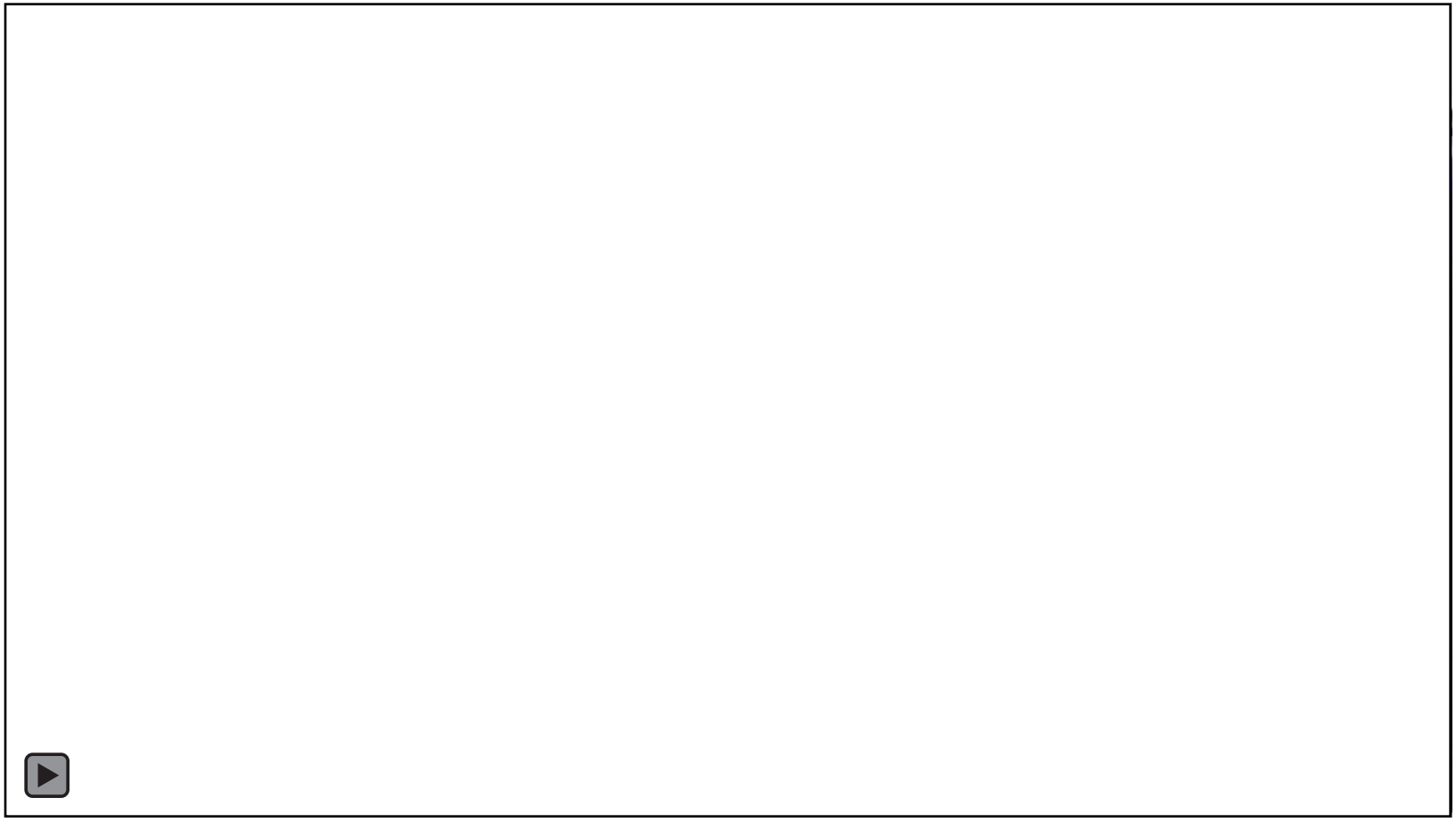
Study 2 – Overview

- The same four listening tasks

- Retrospective recalls
- Stimulated recalls
  - based on eye-traces
- Post-hoc interviews



Study 2 – Stimulated recall procedure



Here I was thinking, like, because they had already mentioned this movie…

Study 2 – Research design

# of times heard	Version 1	Version 8
	example task	example task
1	MC 1	OF 2
	retr. recall	retr. recall
	stim. recall	stim. recall
2	MC 2	OF 1
	retr. recall	retr. recall
	stim. recall	stim. recall
	interview	interview

→ 8 different versions

- 2 participants/version
- 16-18 year-old Austrian school students
- 16 participants overall

Findings – Response processes

construct-relevant	construct-irrelevant
cognitive processes	test-taking strategies
listening strategies	anxiety

Findings – Cognitive processes (Study 2)

(Field, 2013; Vandergrift and Goh, 2012)

lexical search – decoding individual words and phrases

P10: And then I somehow/then he said "highlander".

parsing – bare meaning at clause or sentence level

P14: And then she said "improved your image".

meaning construction – meaning in context

P13: Yes, so she was a difficult girl when she was 7 [...] but they did not say that she had mental problems.

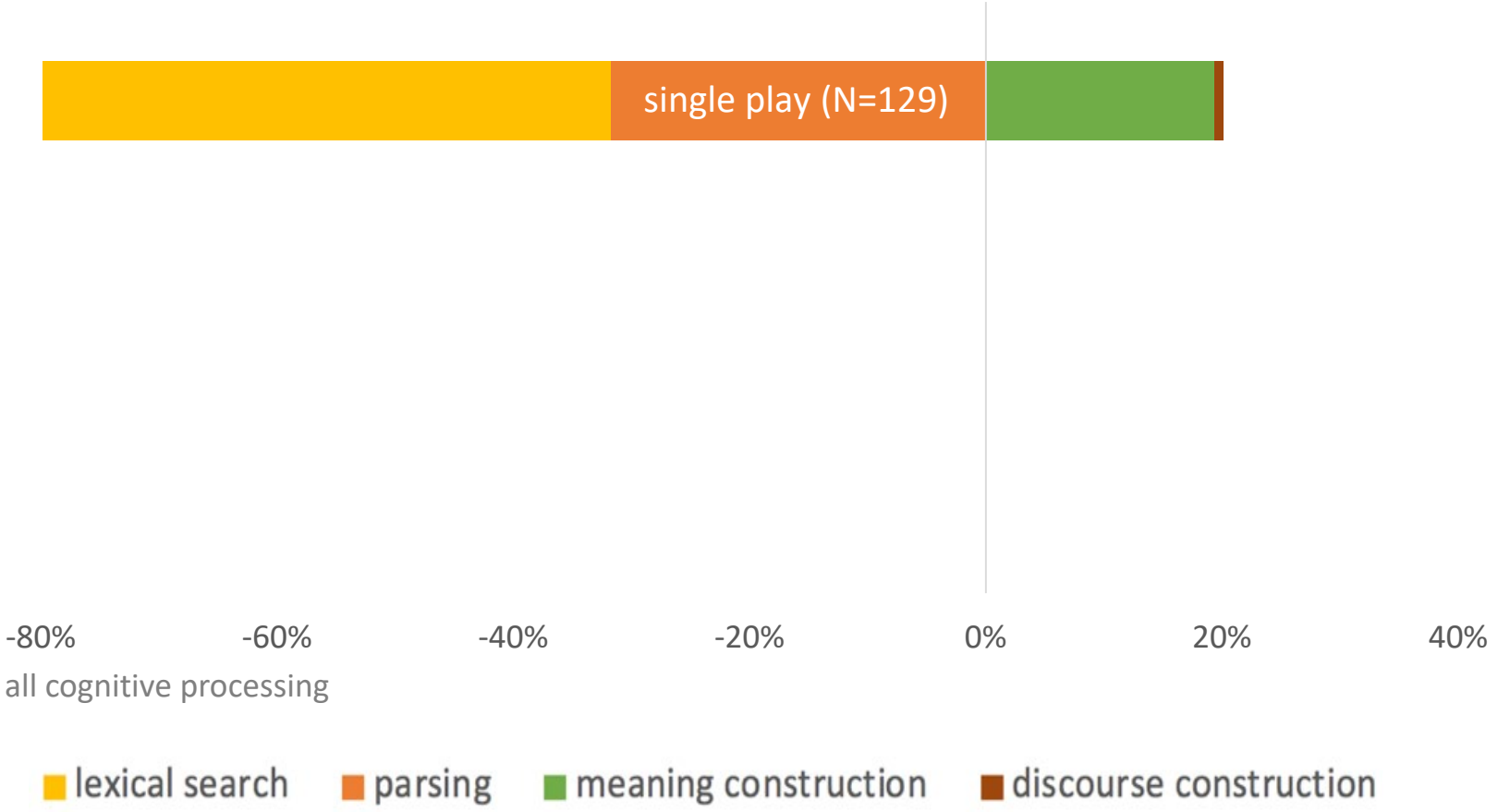
discourse construction – meaning at discourse level

e.g. accurate summary of the listening text

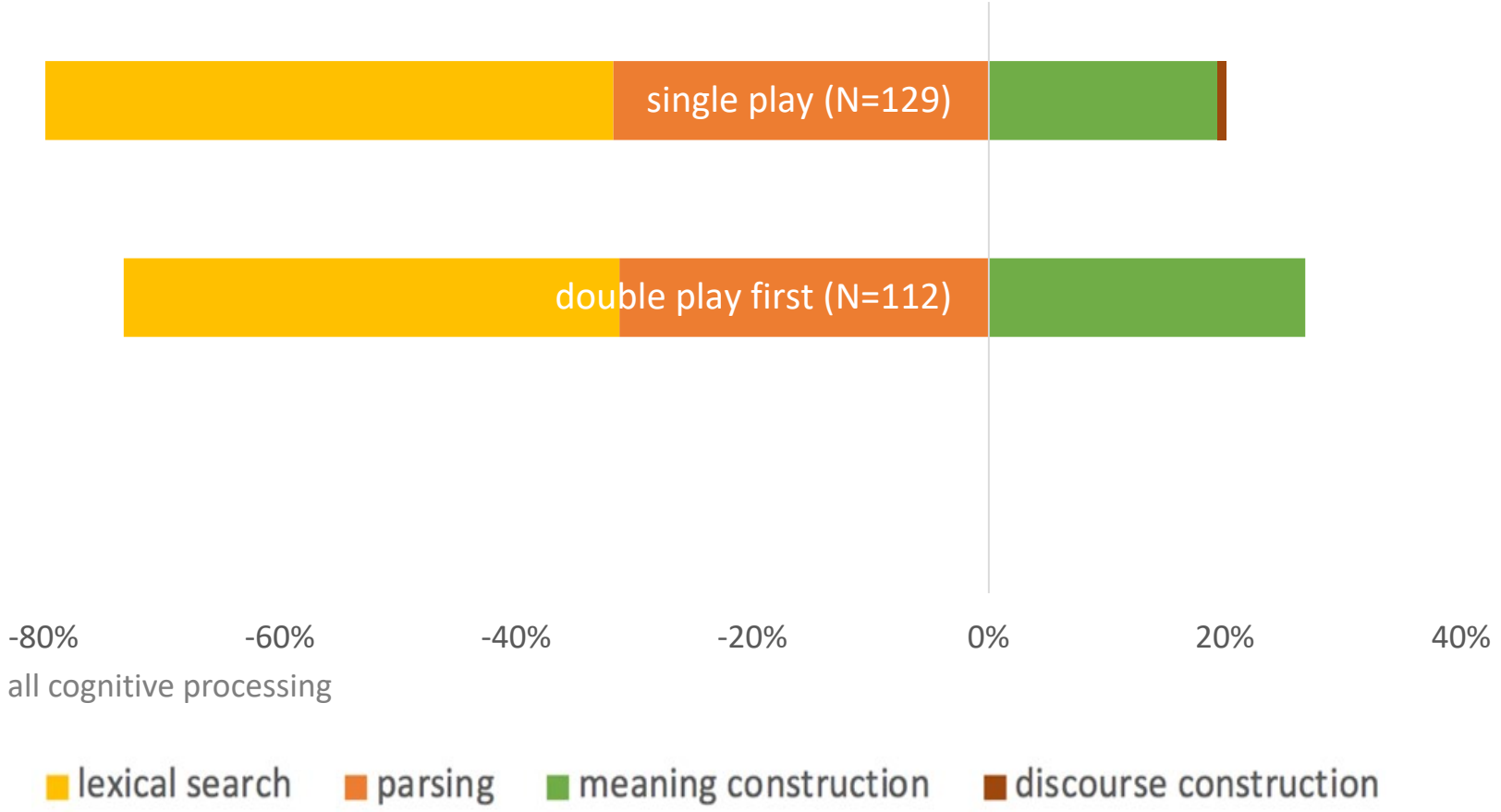
lower-level

higher-level

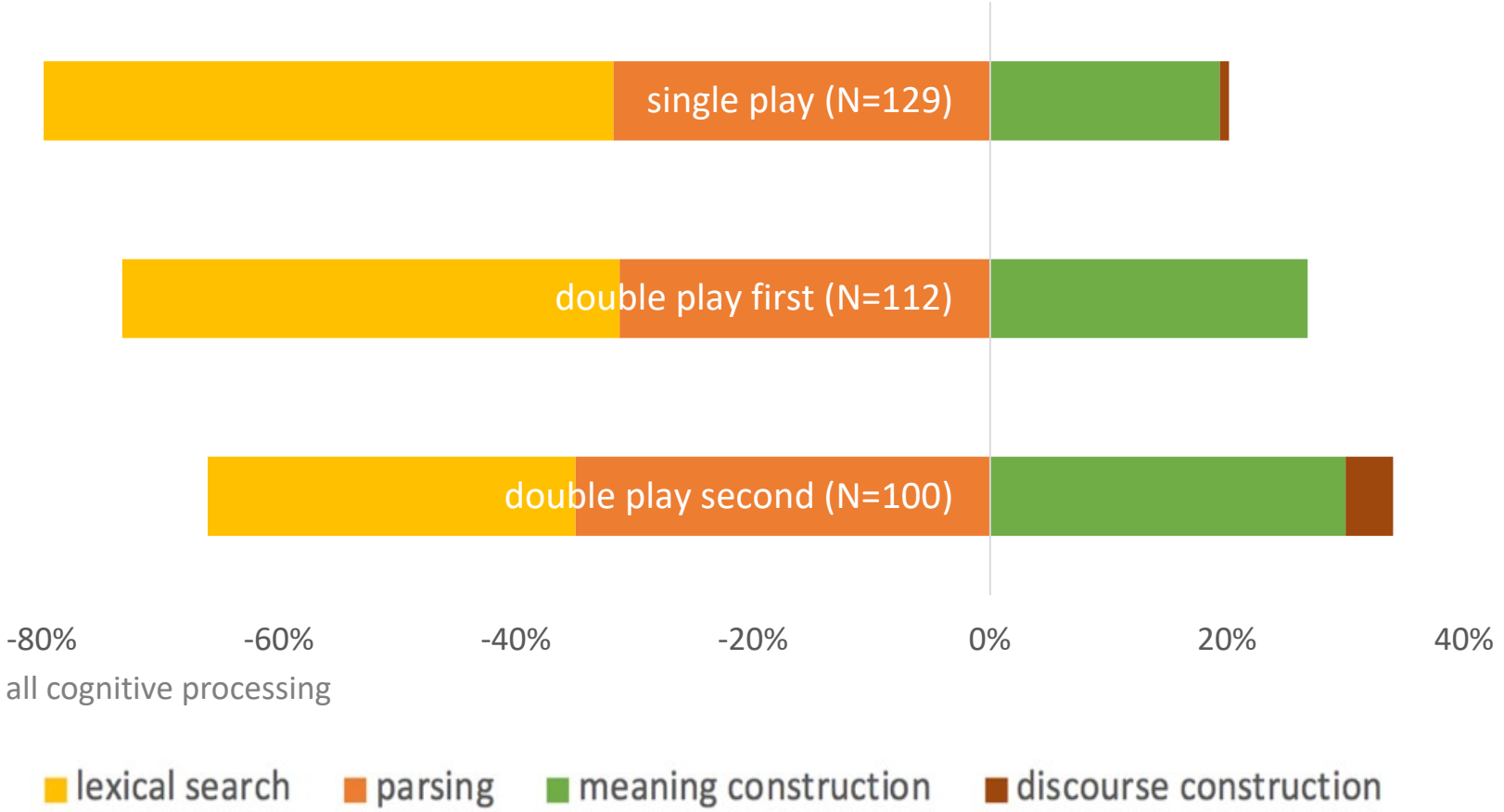
Findings – Cognitive processes (Study 2)



Findings – Cognitive processes (Study 2)



Findings – Cognitive processes (Study 2)



Candidates use more meaning-building processes in double play

Findings – Cognitive processes (Study 2)

P14: So now [in single play] I tried to pay attention to specifics, like, how much this is. [I was] listening more to the details than before.

P04: [During the first play] I am so focused on answering the questions, so [during the second play] I understand things which I don't understand when I answer. […] When I'm not focusing on answering questions I understand more of the context.

Findings – Listening strategies

Study 1

Questionnaire (n=306):

Candidates used more listening strategies in double play

(Wilcoxon signed-rank Z=-4.09, p<.00, r=-0.17)

Study 2

Candidates used a greater variety of listening strategies in double play (N=10) than single play (N=8)

Findings – Listening strategies

P10: Yes, so [when I can hear it twice] I, like, […] think more about it. I kind of let the text get to me, […] like, I engage more with the text. When I hear it only once I have the feeling of, like, being at war with the text and [when I hear it twice] it's much nicer, so that I have the feeling ok now I'm working with the text.

Findings – Test-taking strategies

Study 1

Questionnaire (n=306):

Candidates used more test-taking strategies in single play

(Wilcoxon signed-rank

Z=-1.93, p=.05, r=0.07)

Findings – Test-taking strategies

Study 1

Questionnaire (n=306):

Candidates used more test-taking strategies in single play

(Wilcoxon signed-rank

Z=-1.93, p=.05, r=0.07)

Study 2

test-management strategies

test-dependent and language-dependent

test-wiseness strategies

test-dependent but language-independent

(Cohen, 2011; Doe & Fox, 2011)

Findings – Test-taking strategies

Study 1

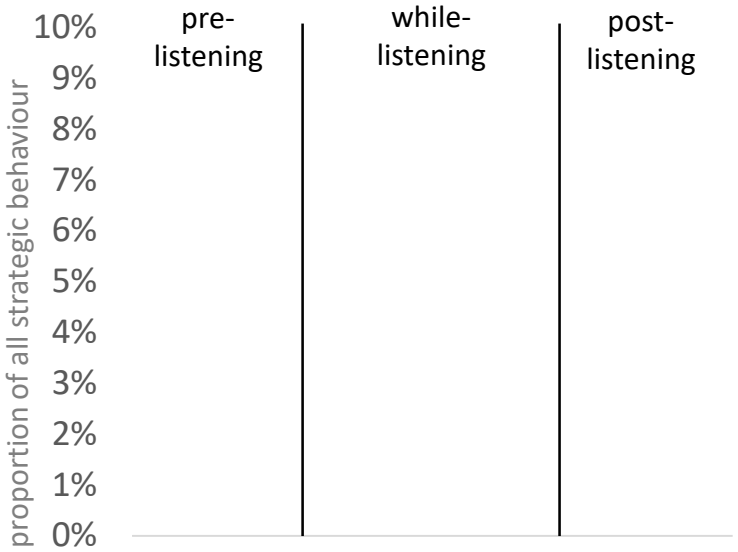
Questionnaire (n=306):

Candidates used more test-taking strategies in single play

(Wilcoxon signed-rank

Z=-1.93, p=.05, r=0.07)

Study 2: test-wiseness



Findings – Test-taking strategies

Study 1

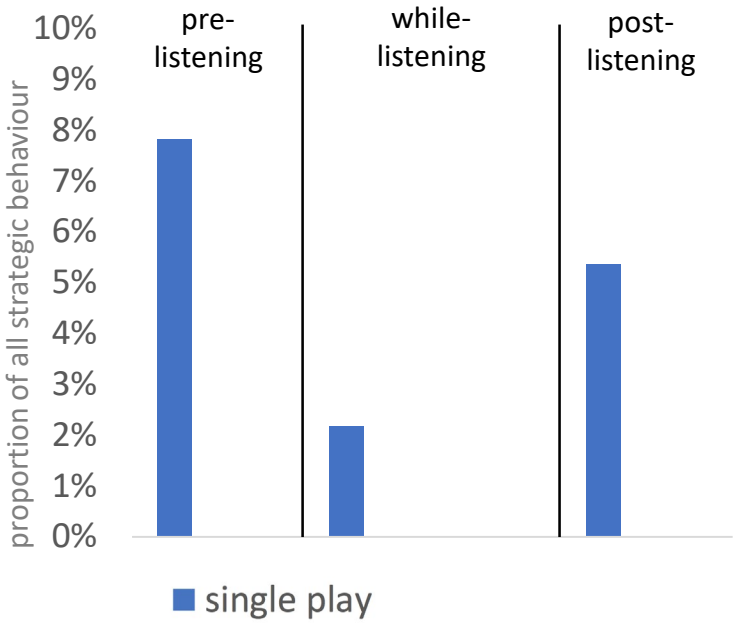
Questionnaire (n=306):

Candidates used more test-taking strategies in single play

(Wilcoxon signed-rank

Z=-1.93, p=.05, r=0.07)

Study 2: test-wiseness



Findings – Test-taking strategies

Study 1

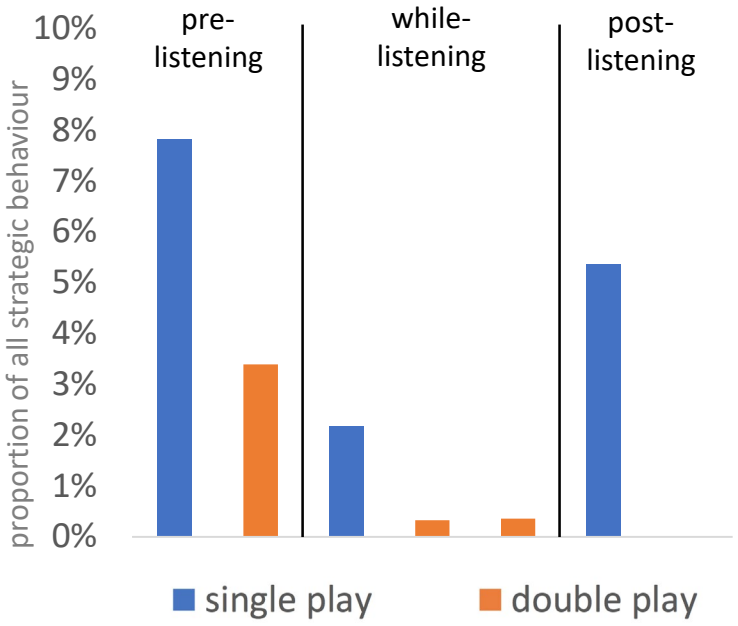
Questionnaire (n=306):

Candidates used more test-taking strategies in single play

(Wilcoxon signed-rank

Z=-1.93, p=.05, r=0.07)

Study 2: test-wiseness



Findings – Test-taking strategies

Study 1

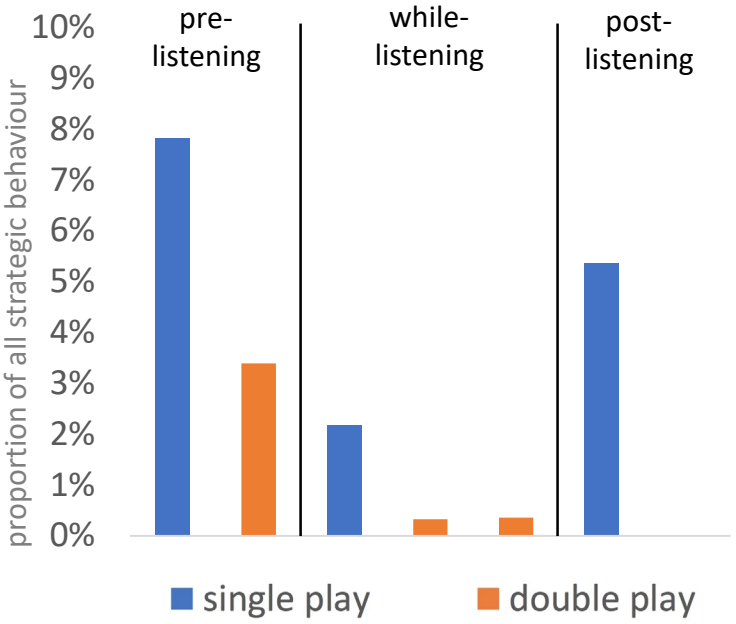
Questionnaire (n=306):

Candidates used more test-taking strategies in single play

(Wilcoxon signed-rank

Z=-1.93, p=.05, r=0.07)

Study 2: test-wiseness



Candidates rely less on test-taking strategies in double play

Findings – Test-taking strategies

P15: Yes, so now [in single play] I looked very closely at the questions, because, like, as I said before, [during double play] I first listen […], but that would not be so wise now, because I won't hear it a second time.

P13: So I think when I know that I will hear it only once […] I try to get all [the answers] during the first play. And if I don't get something then I simply choose whatever I think fits best.

Findings – Anxiety

Study 1

Questionnaire (n=306):

Candidates are more anxious in single play

(Wilcoxon signed-rank

Z=-6.12, p<.00, r=-0.25)

Findings – Anxiety

Study 1

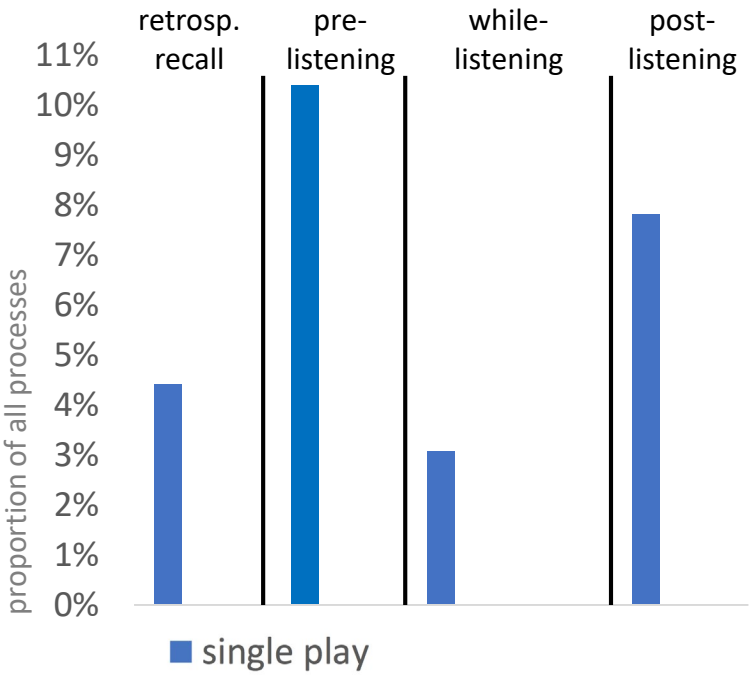
Questionnaire (n=306):

Candidates are more anxious in single play

(Wilcoxon signed-rank

Z=-6.12, p<.00, r=-0.25)

Study 2



Findings – Anxiety

Study 1

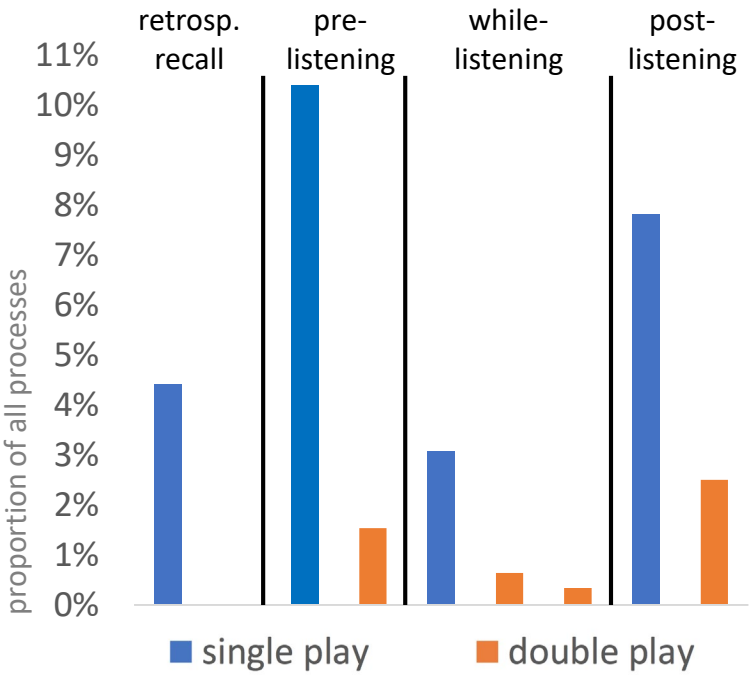
Questionnaire (n=306):

Candidates are more anxious in single play

(Wilcoxon signed-rank

Z=-6.12, p<.00, r=-0.25)

Study 2



Findings – Anxiety

Study 1

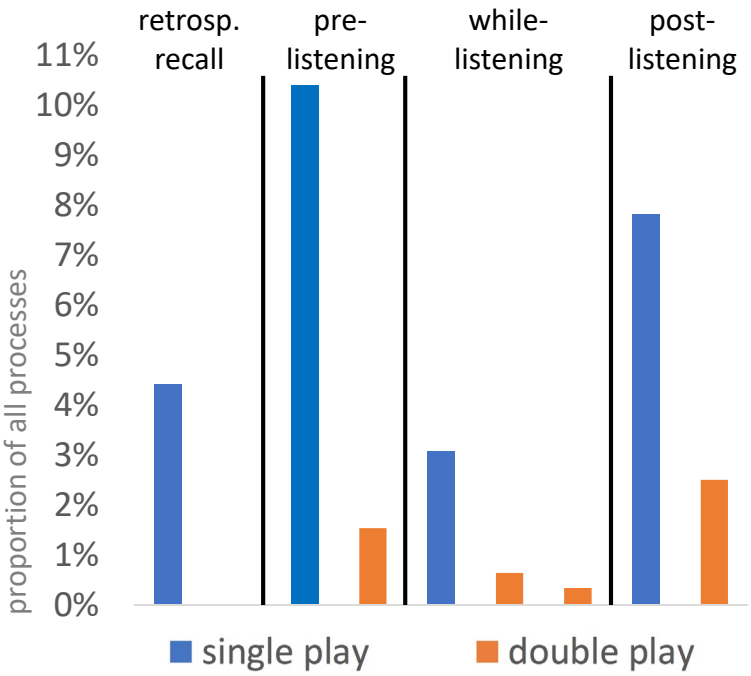
Questionnaire (n=306):

Candidates are more anxious in single play

(Wilcoxon signed-rank

Z=-6.12, p<.00, r=-0.25)

Study 2



Candidates are less anxious in double play

Findings – Anxiety

INT: Ok. Is there anything else you were thinking [while completing the task]?

P16: Yes I felt stressed, because I could only hear it once.

P10: […] It's, like, I can't concentrate when I'm stressed and I constantly think about the stress when I hear it only once.

Implications

Field's (2015) findings confirmed and extended

construct-irrelevant variance	construct-underrepresentation

Implications

Field's (2015) findings confirmed and extended

construct-irrelevant variance	construct-underrepresentation
In single play, test takers…	
• **rely more on test-management strategies**	
• **display more test-wise behaviour**	
• **are markedly more anxious**	
…compared to double play.	

Implications

Field's (2015) findings confirmed and extended

construct-irrelevant variance	construct-underrepresentation
In single play, test takers…	*In single play, test takers…*
• **rely more on test-management strategies**	• **display fewer higher-order cognitive processes**
• **display more test-wise behaviour**	• **use a smaller number of listening strategies**
• **are markedly more anxious**	• **display less listening-strategic behaviour**
…compared to double play.	*…compared to double play.*

Implications

Common argument for single play:

- More tasks can be tested – better for construct representation

However:

- Construct representation better in double play

Implications

Common argument for single play:

- Only one chance in real life – single play more authentic

However:

"The notion that L2 learners must grab a flow of speech on the first try or lose the meaning is valid only for those events where the audio is not repeatable."

(Robin, 2007, p. 110)

Implications

- Increasing role of technology in academic and professional domains

(Graham, Hjorth, & Lehdonvirta, 2017; Sun & Chen, 2016)

- Rise in computer-mediated mobile language learning

(Hubbard, 2017)

➢ Self-paced listening more and more common

- **Double play as the standard convention in general L2 proficiency exams?**

Lancaster University

universität innsbruck

Thank you for your attention!

Questions?

Comments?

franz.holzknecht@uibk.ac.at

@LU_LanguageTest

@LTRGInnsbruck