Exploring Test-taking Processes in a While-listening Performance Test with Question Preview

Zhi Li

March 23, 2018

Language Assessment Research Conference (LARC)

Ames, Iowa

While-listening Performance Tests

- While-listening performance tests & Post-listening tests
  - While-listening: CAEL CE, IELTS
  - Post-listening: TOEFL, CELPIP

While-listening performance tests → Question preview

Question Preview in Listening Tests

- Preview options
  - –Question+Option vs. Question-only vs. Option-only (Koyama, Sun, & Ockey, 2016; Yanagawa & Green, 2008)

- The need for question preview in listening tests
  - \+ Provide a purpose for listening (Buck, 1995; Sherman, 1997)
  - \- May change the way test-takers process input (Hughes, 2003)

- Effects of question preview
  - \+ Benefited low-proficiency test-takers (Sherman, 1997)
  - \+ Benefited advanced learners only (Chang & Read, 2006; Wu, 1998)
  - \+ Benefited the test takers of both levels (Koyama et al., 2016)

Listening Comprehension

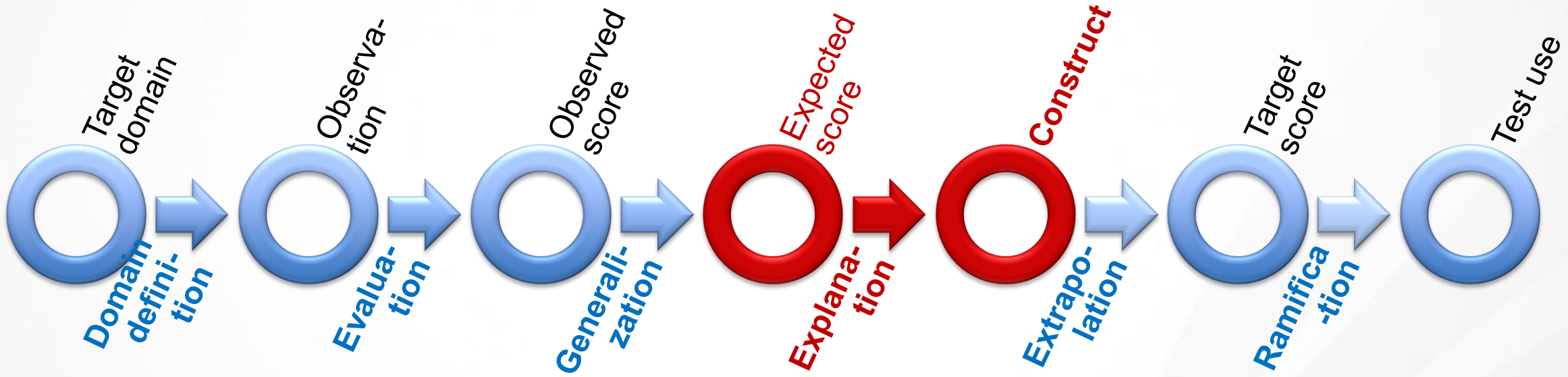
- Conceptualization of listening comprehension
  - –Subskill-based
    - •Listening for local information, comprehending global information, making inferences
  - –Strategy-based
    - •Cognitive strategies & metacognitive strategies
  - –Cognitive process-based
    - •Bottom-up & top-down processing
    - •Controlled processes & Automatic processes (Field, 2013; Green, 2017)
    - •Automaticity in second language processing (Segalowitz,2008)

Responding Processes in While-listening Tests

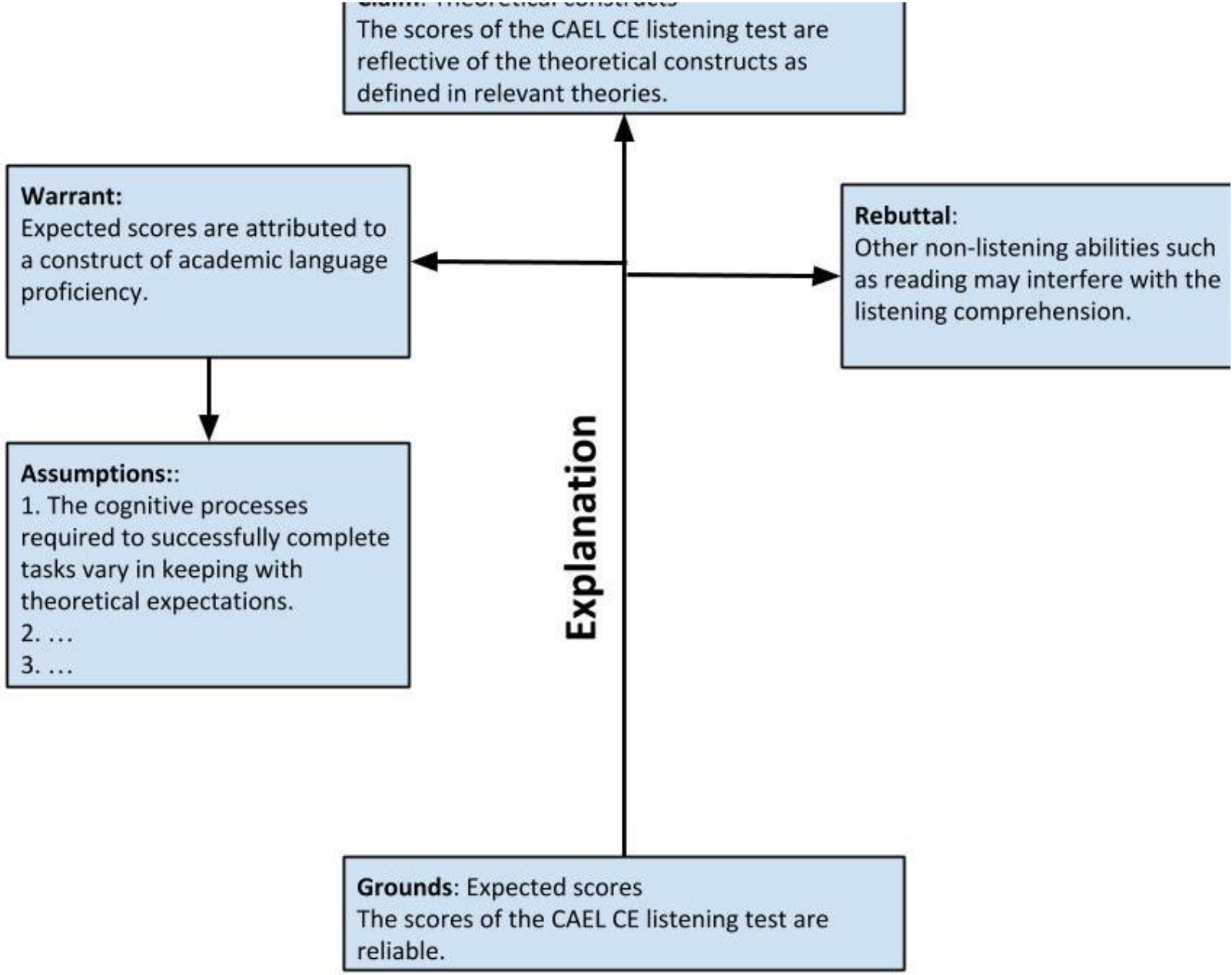
- Field (2013, p. 106-107)
  - *The importance of automaticity in all these processes cannot be overstated. … If a basic operation like matching a set of speech sounds to a word requires an effort of attention, it imposes demands upon a listener's working memory that can preclude other operations. By contrast, when the mapping from word to word senses is highly automatic, working memory resources are freed for higher-level processes such as making inferences, interpreting the speaker's intentions, recognising a line of argument and so on.*

- Lots of studies on listening strategies, but few on responding processes

Argument-based Approach to Validation



Explanation Inference



Research Questions

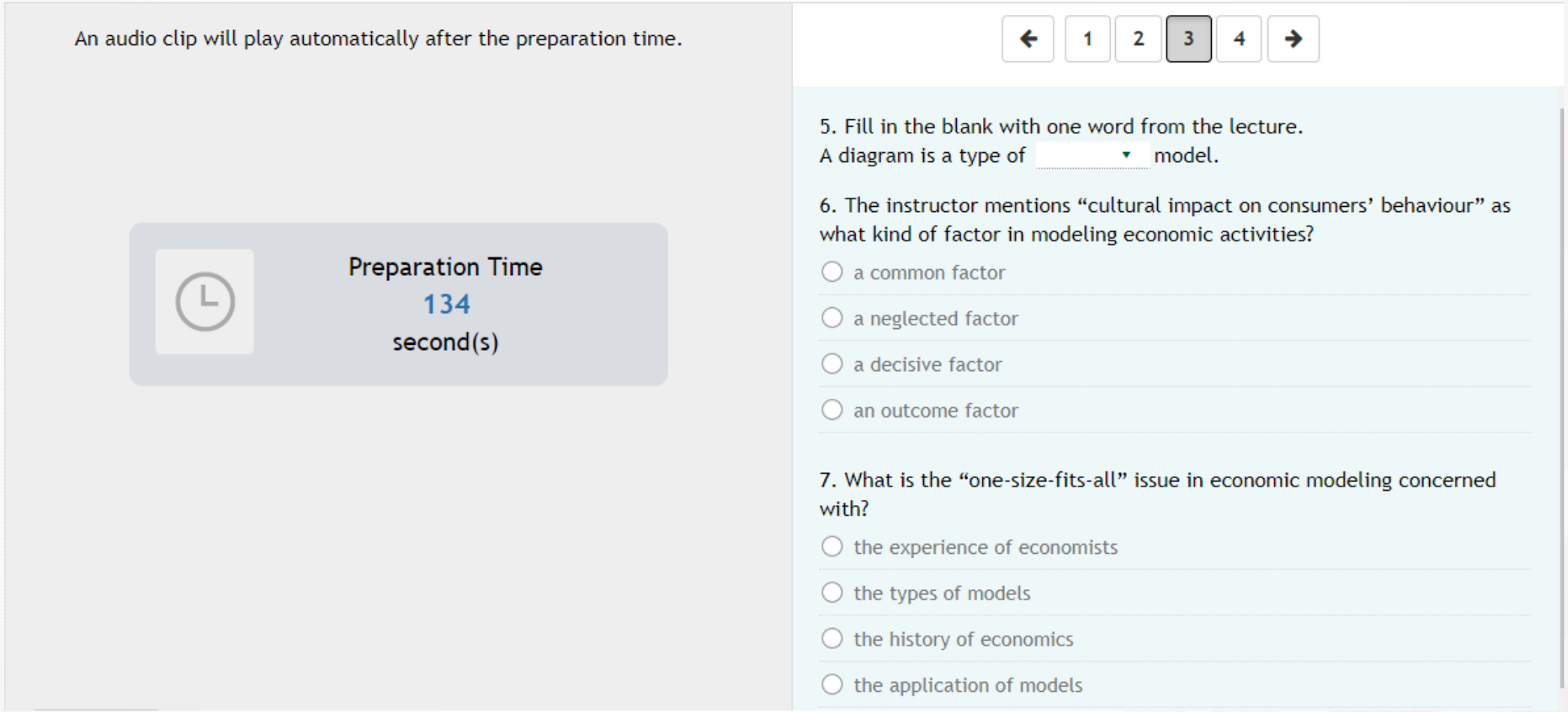
1. To what extent do test takers of different listening proficiency levels differ in their question-preview behavior?

2. To what extent do test takers of different listening proficiency levels differ in their responding processes?

The CAEL CE Listening Test

- The Canadian Academic English Language (CAEL) Test, Computer Edition (CE)
  - An integrated and topic-based test of English for academic purposes (https://www.cael.ca/)
  - Five Parts, computer-delivered:
    - Speaking, Integrated Reading, Integrated Listening, Academic Unit A, and Academic Unit B
  - While-listening performance test
    - One short and three long listening testlets
    - Long listening testlets = mini-lecture on an academic topic
    - Item formats include standard MCQ, in-line choice, fill in the blanks, matching task

The CAEL CE Listening Test – Sample Interface



Note: This is a screenshot of an example listening test.

Paragon Testing Enterprises

The CAEL CE Listening Test – The Study Testlet

- Subskills
  - –Comprehending local information (6 Items)
  - –Comprehending global information (3 Items)
  - –Making inferences (2 Items)
- Item format
  - –MCQ with 4 options (in-line choice & regular layout)

Topic	Duration of Question Preview	Duration of Lecture	Duration of Post-lecture Time	Item Configuration
Psychology	2 min. 30 sec.	5 min. 48 sec.	2 min.	P1: 1, 2, 3; P2: 4, 5; P3: 6, 7, 8; P4: 9, 10, 11

The CAEL CE Listening Test – The Study Sample

104 Participants (after excluding 10 outliers) recruited for a pilot test

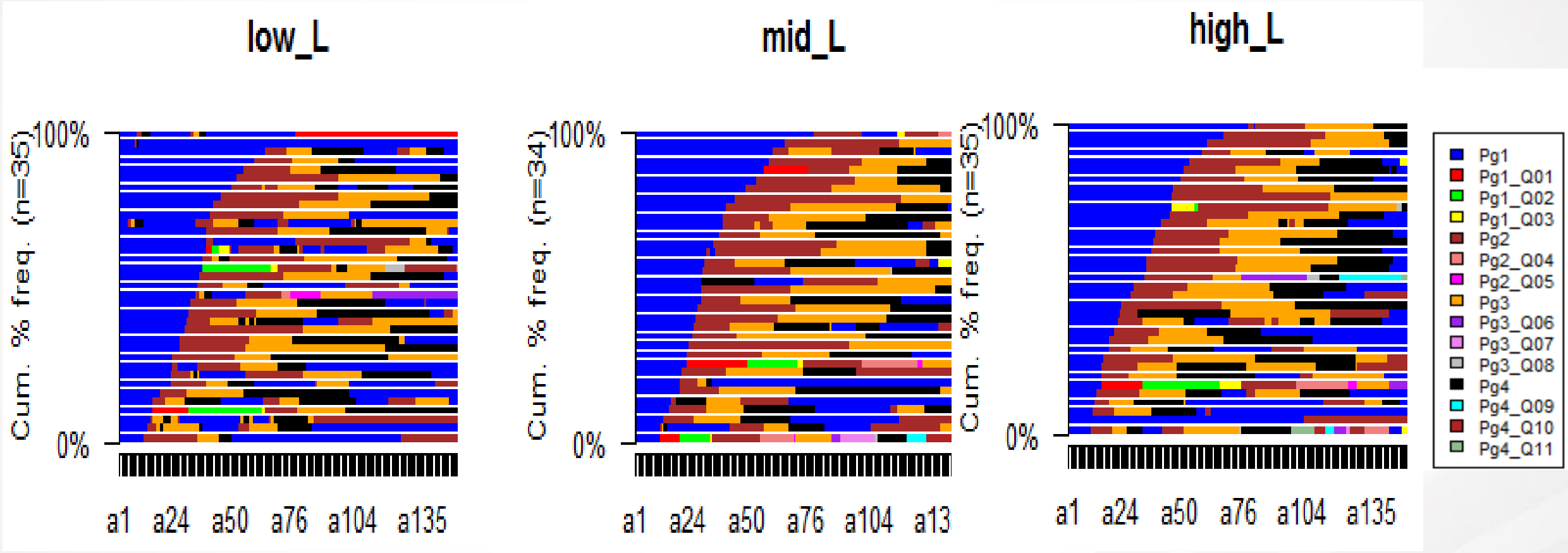
- Low (n=35): Average 27.8 (out of 100), SD 7.5
- Mid (n=34): Average 50.3 (out of 100), SD 7.5
- High (n=35): Average 81.2 (out of 100), SD 10.0

Data Collection and Analysis

- Data
  - Test score data
  - Timestamped behavior log data

- Analysis
  - State Sequence Analysis using R Package *TraMineR* (Gabadinho et al., 2011)
  - Visual examination of question-previewing behaviors and responding processes
  - Non-parametric tests for the comparison of time allotments in the question-previewing stage
  - Non-parametric tests for the comparison of time allotments in the lecture stage

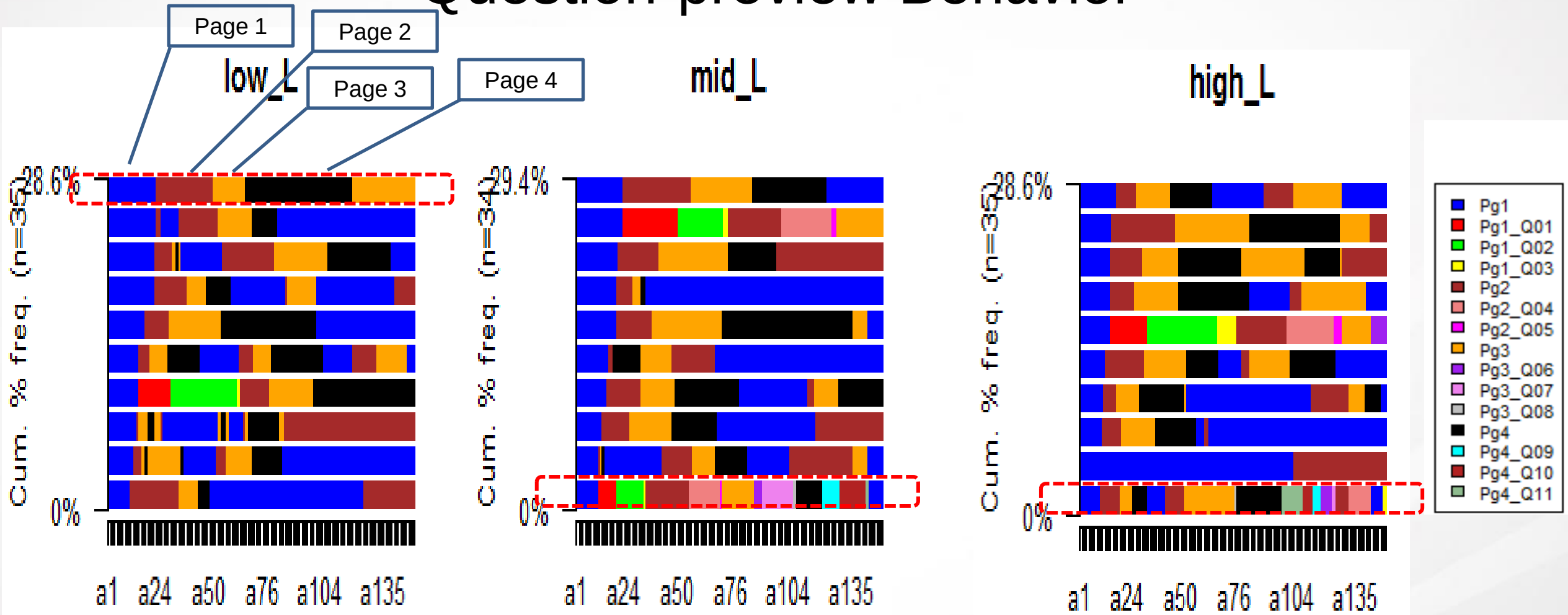
Question-preview Behavior



Sequence frequency plot for each group during the question-preview stage

Paragon TESTING ENTERPRISES

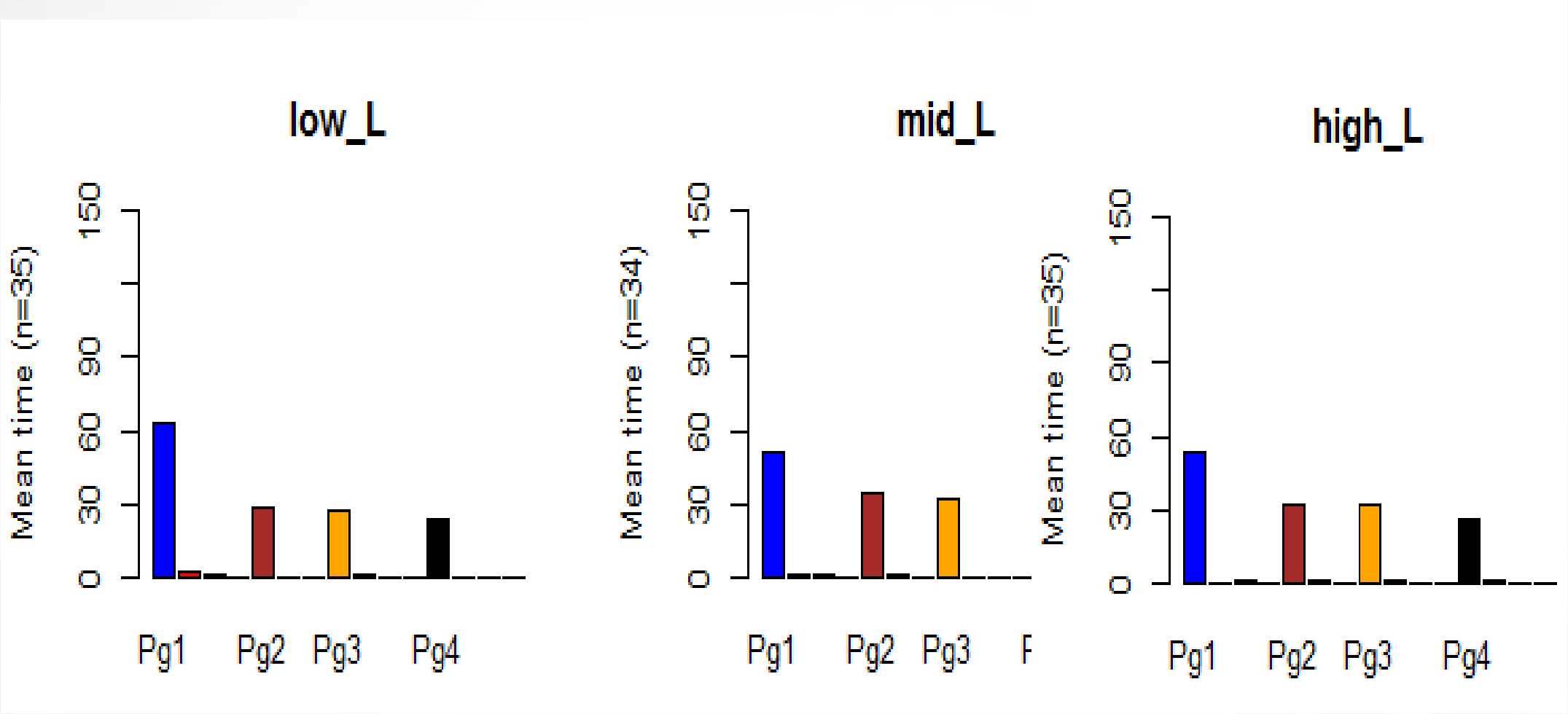
Question-preview Behavior



Sequence frequency plot for the first 10 cases in each group during the question-preview stage

Paragon TESTING ENTERPRISES

Question-preview Behavior – Time Allotment



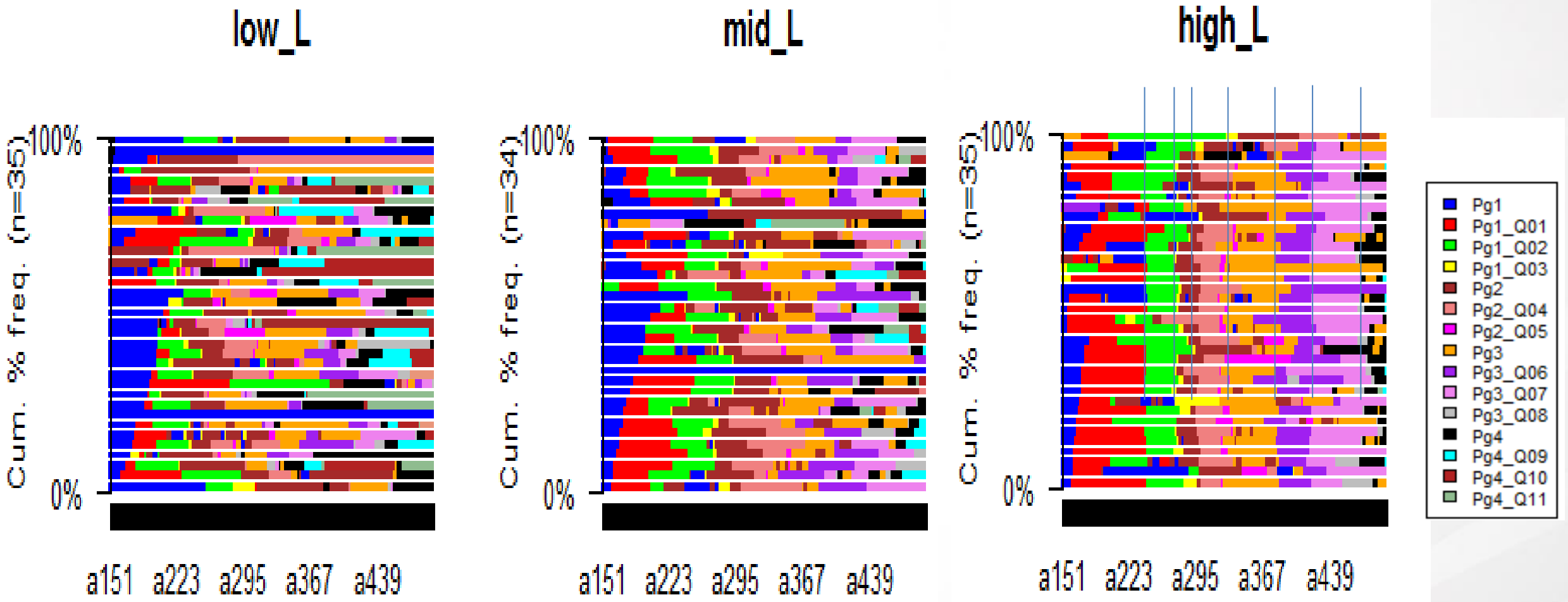
Average time allotment during the question-preview stage

Paragon TESTING ENTERPRISES

Summary: Question-preview Behavior

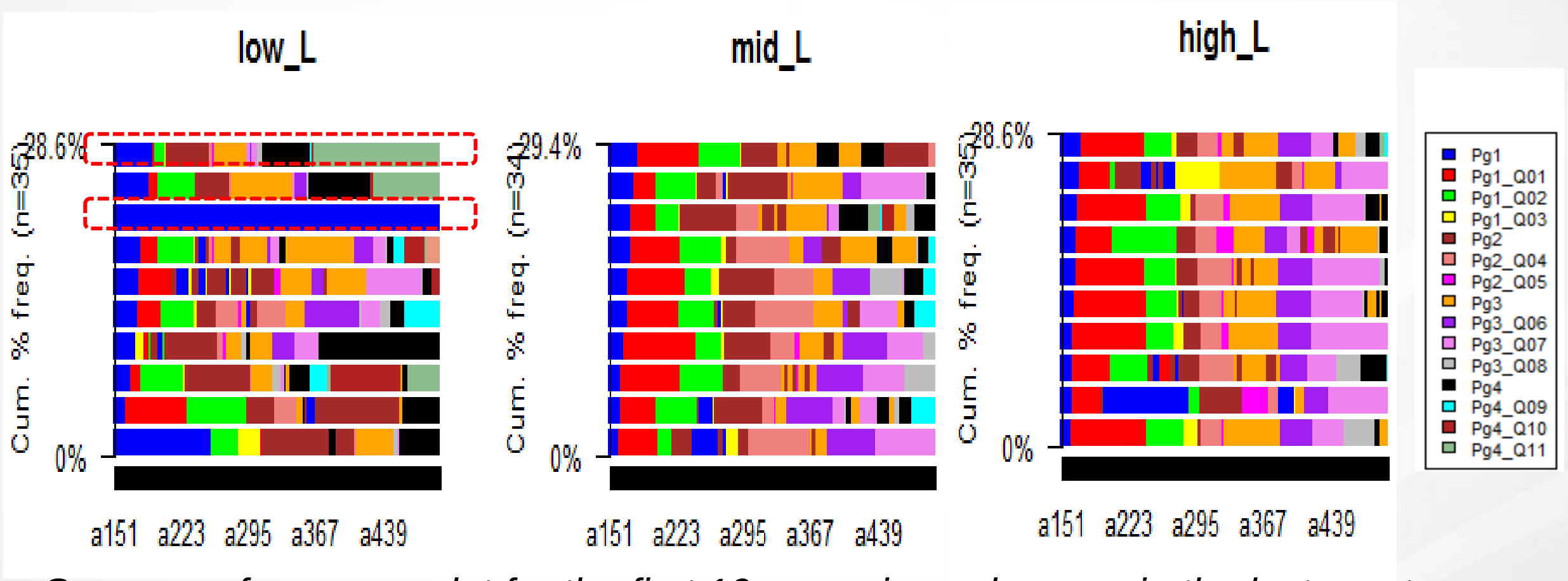
- All the test takers could finish previewing the questions
- Different preview approaches are observed.
- There were no significant differences in terms of time allotment among the three proficiency groups

Responding Processes



Sequence frequency plot for the ALL cases in each group in the lecture stage

Responding Processes



Sequence frequency plot for the first 10 cases in each group in the lecture stage

Responding Processes – Time Allotments

Item	Subskill	Item Format	Difficulty	Discrimi-nation	Difference in Time Allotments
2	Global	In-line Choice	0.43	0.54	H > M&L
3	Local	In-line Choice	0.51	0.27	H > M&L
5	Global	MCQ	0.30	0.47	H < M&L
7	Local	MCQ	0.63	0.37	H > M&L
8	Local	In-line Choice	0.26	0.33	H > M&L

Note: H: High proficiency group; M = Medium proficiency group; L = Low proficiency group

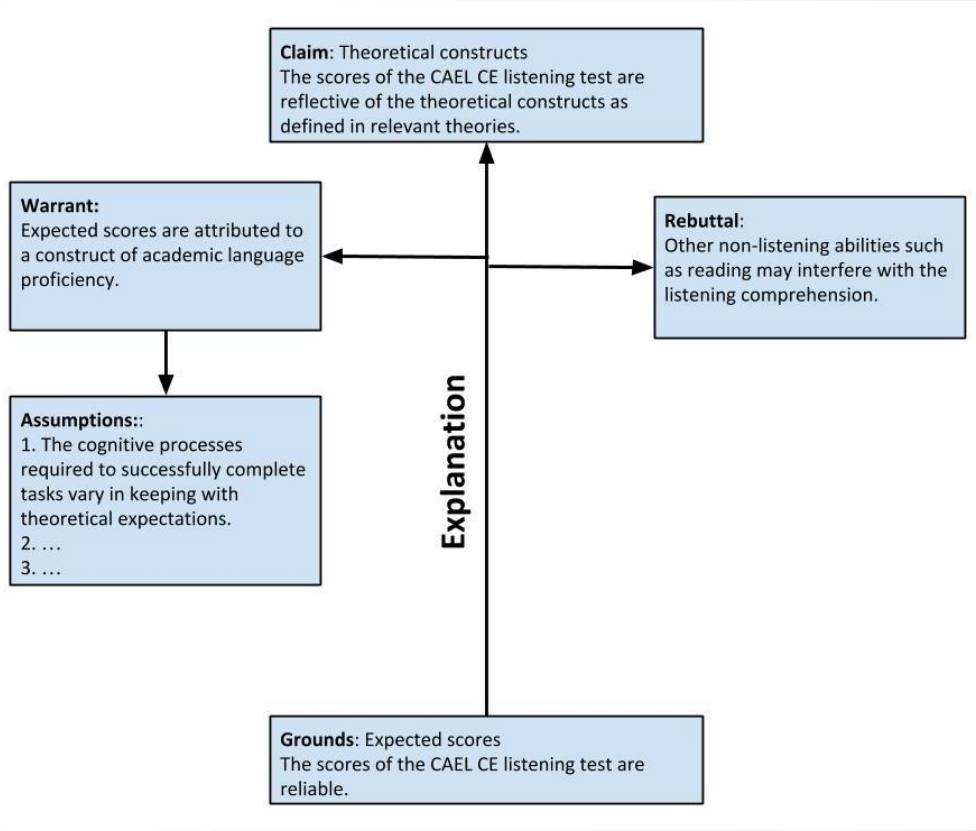
Summary: Responding Processes

- Noticeably different responding patterns or progression patterns were observed among the three proficiency groups
- High-performing group seemed to be able to follow closely with the lecture and respond to items in a more timely manner.
- There were some differences in the time allotment on individual items in the test. More investigation is needed to find out what caused these differences.

Explanation Inference

- The results in this study lend support or backing to the assumption (RQ2), while providing evidence to partially refute the rebuttal (RQ1).
- Implications

- More studies are needed to study other relevant assumptions for this inference.



TESTING ENTERPRISES Paragon

Limitations & Future Studies

- The participants
  - Limited demographic information
  - Possible variations in the motivation levels in this pilot test

- The testlet
  - Single testlet -> limited generalizability
  - Small N-size

Thank you!
Questions & Comments?

Zhi Li (zli@paragontesting.ca)

Paragon Testing Enterprises, Inc. Vancouver, Canada

Selected References

- Aryadoust, V. (2011). Application of the fusion model to while-listening performance tests. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 15(2), 2–9.
- Berne, J. E. (1995). How does varying pre-listening activities affect second language listening comprehension? *Hispania*, 78(2), 316–329.
- Goh, C. C. . (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75.
- Gabadinho, A., Ritschard, G., Muller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.
- Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, 34(2), 165–182. http://doi.org/10.1016/j.system.2005.11.001
- Yanagawa, K., & Green, A. (2008). To show or not to show: The effects of item stems and answer options on performance on a multiple-choice listening comprehension test. *System*, 36(1), 107–122. http://doi.org/10.1016/j.system.2007.12.003
- Koyama, D., Sun, A., & Ockey, G. (2016). The effects of item preview on video-based multiple-choice listening assessments. *Language Learning & Technology*, *20*(1), 148–165. Retrieved from http://lib.dr.iastate.edu/engl_pubs/73
- Segalowitz, N. (2008). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 382–408). Oxford, UK: Blackwell Publishing Ltd. http://doi.org/10.1002/9780470756492.ch13
- Sherman, J. (1997). The effect of question preview in listening comprehension tests. *Language Testing*, 14(2), 185–213.
- Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes*, 10(2), 89–101.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40(3), 191–210.