Statistical analysis in the language sciences

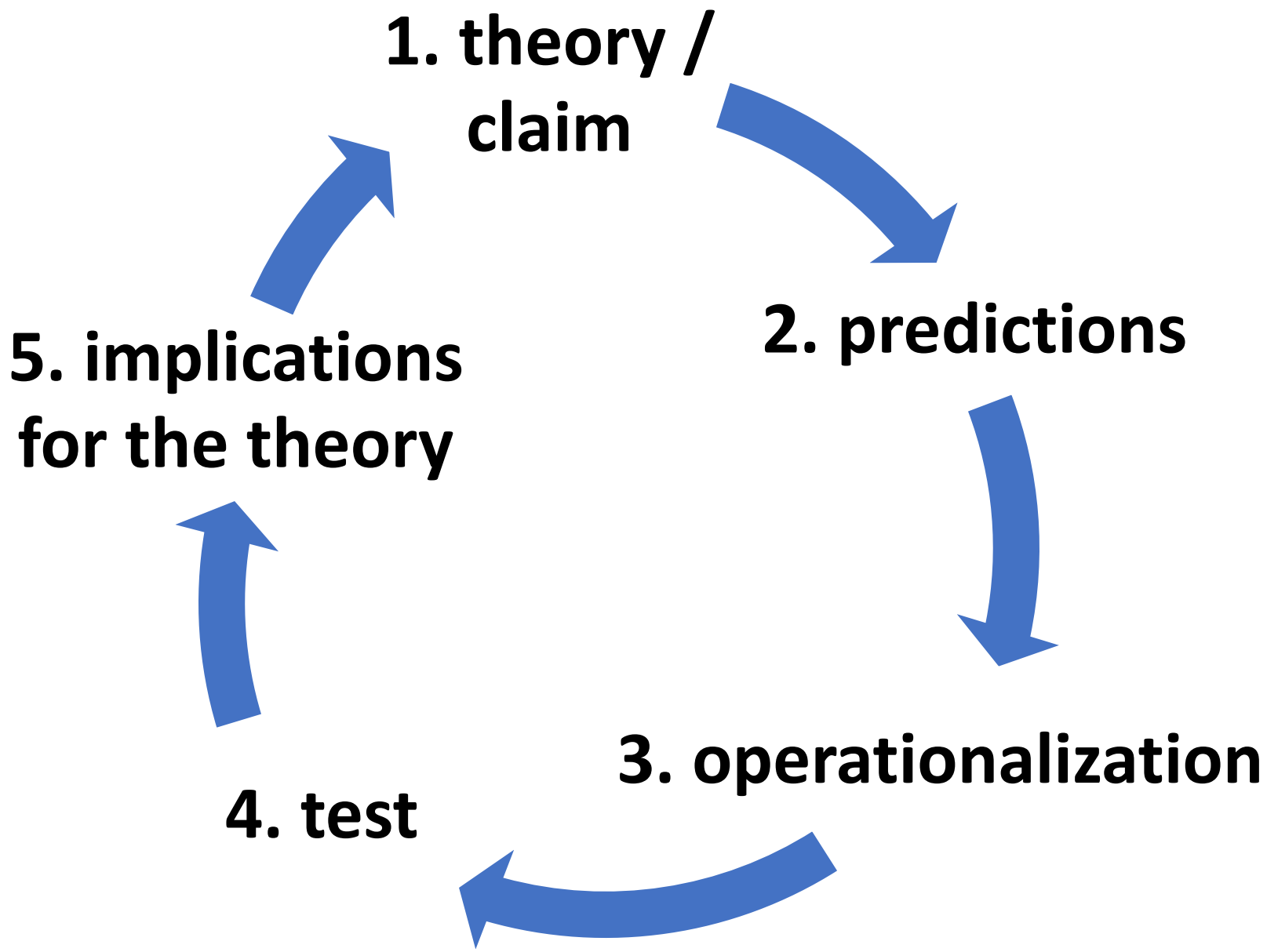
Joost van de Weijer
Laboratory of the Humanities /
Department of linguistics
Lund University

Overview

- Structure of a typical experimental study
- Purpose of a statistical analysis
- Relation between the setup of the study and statistical analysis
- Shaping the study to fit the statistical analysis
- Five example studies

Structure of an experimental study

- There is an existing theory.
- Predictions are made what will happen if the theory is correct.
- The predictions are being translated into a manageable and measurable design
- Data are collected and analyzed
- If the results are consistent with the predictions the theory is accepted. If not, the theory is rejected, and a new theory is formulated.


1. theory / claim
2. predictions
3. operationalization
4. test
5. implications for the theory

- The operationalization of the predictions in experimental research often consists of manipulating one or more variable and measuring the effect that this manipulation has on another variable.
- The manipulated variables are the experiment's **independent variables** (also factors, or predictors)
- The measured variable is the experiment's **dependent variable** (also outcome variable)
- Variables can be **numerical** (e.g., reaction time, age, proficiency score) or **categorical** (e.g., language, grammatical construction)
- Categorical variables can be **ordered** (e.g., small-medium-large)

An example

- My theory is that tone perception (as in Chinese) is easier for speakers of a word-accent language (as Swedish) than for speakers of a language without word accents (as English).
- I create a set of Chinese words with different tones, and ask a group of five Swedish speakers and five English speakers to learn these words.
- After learning, their knowledge is tested through a perception experiment.

Purpose of the statistical analysis

- Experimental studies are commonly based on relatively small groups of subjects (**samples**)
- This implies that there is uncertainty to the outcome of the experiment: if the exact same experiment is done with a different sample, the results will most likely be not exactly the same (**sampling error**)
- The outcome of the statistical analysis provides an indication (***p***) of how likely it is that observed differences are due to sampling error or not.
- If the indication that the observed differences are *not* due to sampling error is strong, then the result is said to be **statistically significant**.

Shaping the study to fit the statistical analysis

- Doing a statistical analysis is often a requirement for being published.
- There is a large choice of different analyses. Some analyses are easy to apply, others are more complicated. Some are easy to interpret, others are difficult to interpret.
- All analyses have their limitations, i.e., they can only be applied meaningfully if certain requirements apply. The requirements may be related to the kinds of variables (numerical/categorical), the numbers of observations, the shape of the distribution of the variables.
- Therefore, it is necessary to be aware of the possibilities and the limitations (strengths and weaknesses) of different kinds of analyses when setting up the study.

Strengths and weaknesses

- Different tests require different data sets. Many frequently used statistical analyses (regression, ANOVA, t-test) are designed for a single dependent variable with one or more independent variables.
- Statistical tests are not good at handling very small data sets, nor at very large data sets
- Statistical tests are based on the idea that there is variation in the data (if there is no variation, there is no uncertainty, and therefore no need for a comparison)
- Statistical tests for numerical data tend to be more well developed than tests for categorical data (categorical data always need to be transformed into numbers somehow)

Implications

- Design the study such that the entire data set can be analyzed within the framework of a single statistical analysis.
- Reduce the number of independent variables
  - The risk that independent variables are correlated increases when there are many of them. Therefore designs with many independent variables often provide results that are more difficult to interpret than designs with few independent variables.
  - The more independent variables there are, the more cases you need.
- Avoid independent variables that are correlated with each other.
- Choose variables so that you know that there will be sufficient amount of variation in the data.
  - Tasks should not be too easy or too difficult.
  - By their very nature, some variables (e.g., response times) show more variability than other variables (e.g., yes/no responses)

Elements of a statistical analysis

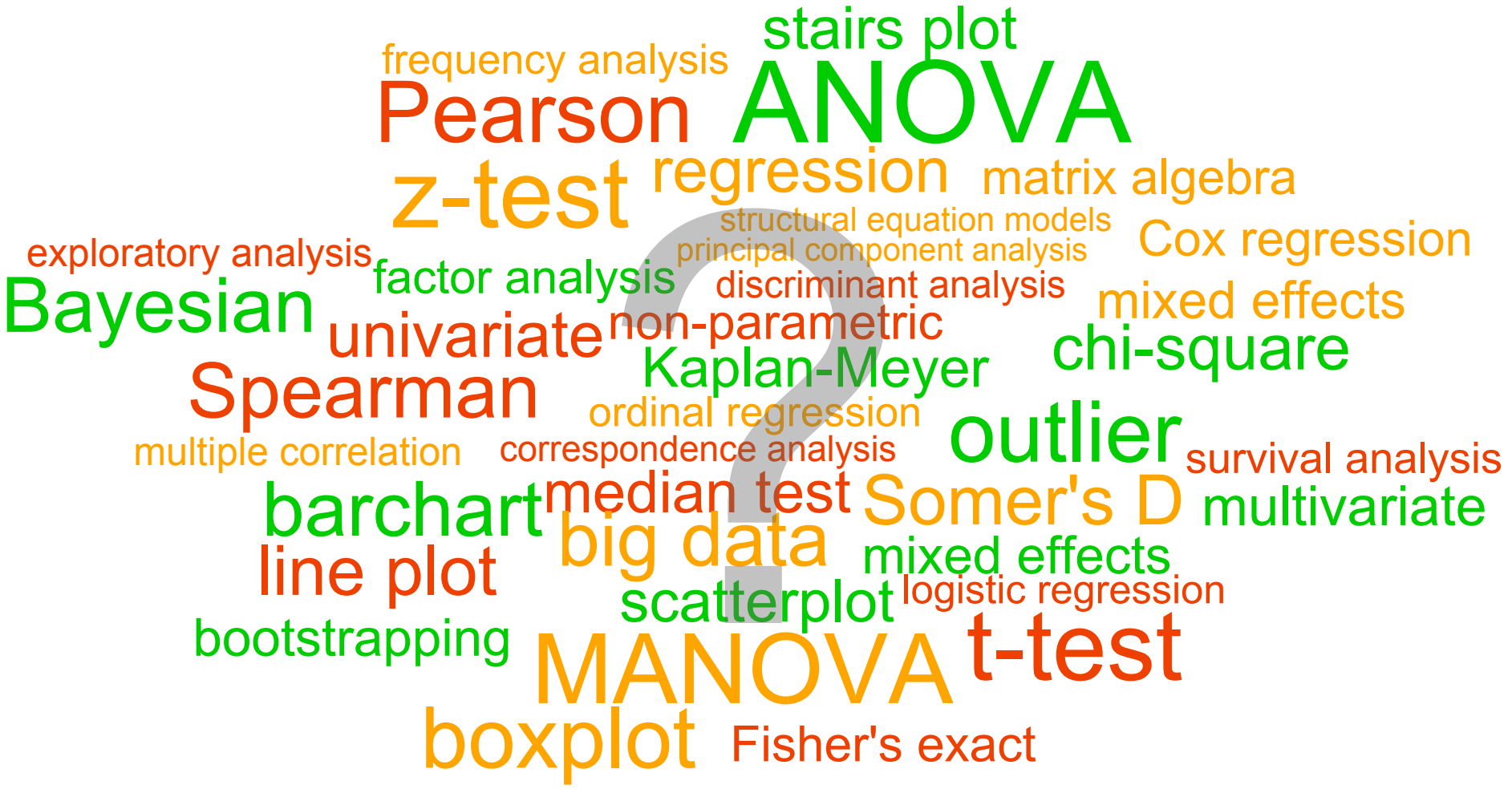
Exploratory analysis

- Purposes
  - Is there enough variation in the data?
  - Are there extreme values (outliers)?
  - Is there correlation between the independent variables?
  - Can I formulate preliminary expectations what the results look like?
  - Are there errors in the datafile?
- Tools
  - Data visualization (scatter plots, boxplots)
  - Summary statistics (mean, median, range)

Descriptive and inferential statistics

- Descriptive statistics consists of providing summary statistics and expected values for the dependent variable for different values of the independent variable(s).
- Common summary statistics (mean) may not be the best estimates of the expected values, especially if there is unbalance in the dataset.
- Instead, the expected values are derived from the analysis.
- The analysis also provides estimates of the effects of the independent variables(s), and whether or not the effects are significant.

Which statistical test?



Recommended analysis: Regression

- It focuses on the effect that one or more independent variables have on a single dependent variable.
- Does the same as ANOVA but also provides estimates of expected values and of effects, and therefore used to describe the results.
- It can be used for the analysis of experimental data, but also correlational data (which are not based on variable manipulation, and therefore often more unbalanced than experimental data), sometimes also for corpus research.
- Variables (dependent and independent) included in a regression analysis can be numerical and/or categorical.

Example studies

Example study 1: Heart with Manuel Oliva

Background and questions

- University literature is often in English, even in non-English speaking countries.
- Illustrations to the text may aid in a better understanding of the contents of the text.
- How should the information be presented? Illustration integrated with the text or separate?
- Do illustrations help readers with lower degrees of English proficiency more than readers with better English proficiency?

Study design

- Swedish native speakers (university students, not in medicine) were shown different versions of a textbook excerpt about the functioning of the human heart.
- The excerpt consisted of text accompanied by a picture of the heart. The text was either complicated or simple, and could be integrated with the picture or not.
- Before the test, their English proficiency was measured.
- After the test, their understanding of the text was measured.
- During the test, their eye movements were tracked (not yet analyzed).

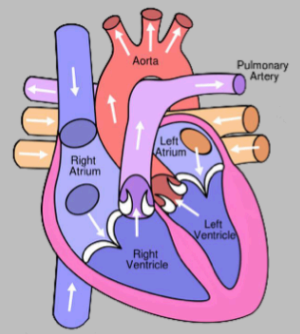
split

Integrated

Simple

Blood Flow in the Human Heart

De-oxygenated blood returns from the body to the heart and is collected in the right atrium. When the ventricles relax, the blood from the right atrium flows into the right ventricle. The ventricles contract and blood is pumped from the right ventricle into the pulmonary artery. The pulmonary artery supplies the lungs, where carbon dioxide is exchanged for oxygen. Oxygenated blood from the lungs is returned into the left atrium. At the same time, blood from the left atrium flows into the left ventricle. Blood is then pumped from the left ventricle into the aorta and back to the body.



Blood Flow in the Human Heart

7. Blood is then pumped from the left ventricle into the aorta and back to the body.

3. The ventricles contract and blood is pumped from the right ventricle into the pulmonary artery.

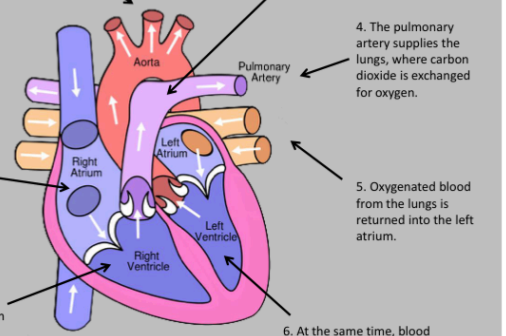
4. The pulmonary artery supplies the lungs, where carbon dioxide is exchanged for oxygen.

1. De-oxygenated blood returns from the body to the heart and is collected in the right atrium.

5. Oxygenated blood from the lungs is returned into the left atrium.

2. When the ventricles relax, the blood from the right atrium flows into the right ventricle.

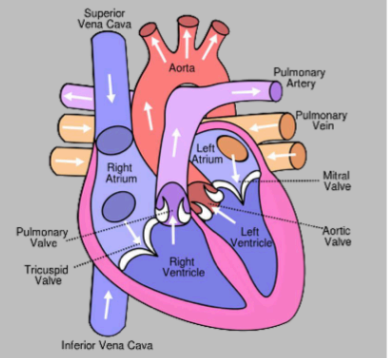
6. At the same time, blood from the left atrium flows into the left ventricle.



Complex

Blood Flow in the Human Heart

De-oxygenated blood returns from the body to the right heart through the superior and inferior vena cava and is collected in the right atrium. When the ventricles relax, the blood from the right atrium flows, through the tricuspid valve, into the right ventricle. The ventricles contract and blood is pumped from the right ventricle, through the pulmonary valve, into the pulmonary artery. The pulmonary artery supplies the lungs, where carbon dioxide is exchanged for oxygen in a passive process of diffusion. In the left heart, oxygenated blood from the lungs is returned via the pulmonary vein into the left atrium. At the same time, blood from the left atrium flows into the left ventricle through the mitral valve. Blood is then pumped, through the aortic valve, from the left ventricle into the aorta and back to the body.



Blood Flow in the Human Heart

7. Blood is then pumped, through the aortic valve, from the left ventricle into the aorta and back to the body.

3. The ventricles contract and blood is pumped from the right ventricle, through the pulmonary valve, into the pulmonary artery.

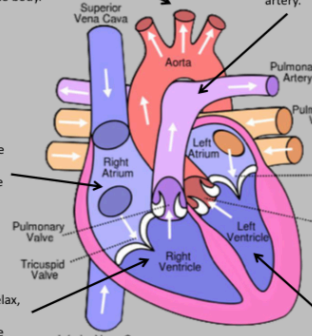
4. The pulmonary artery supplies the lungs, where carbon dioxide is exchanged for oxygen in a passive process of diffusion.

1. De-oxygenated blood returns from the body to the right heart through the superior and inferior vena cava and is collected in the right atrium.

5. In the left heart, oxygenated blood from the lungs is returned via the pulmonary vein into the left atrium.

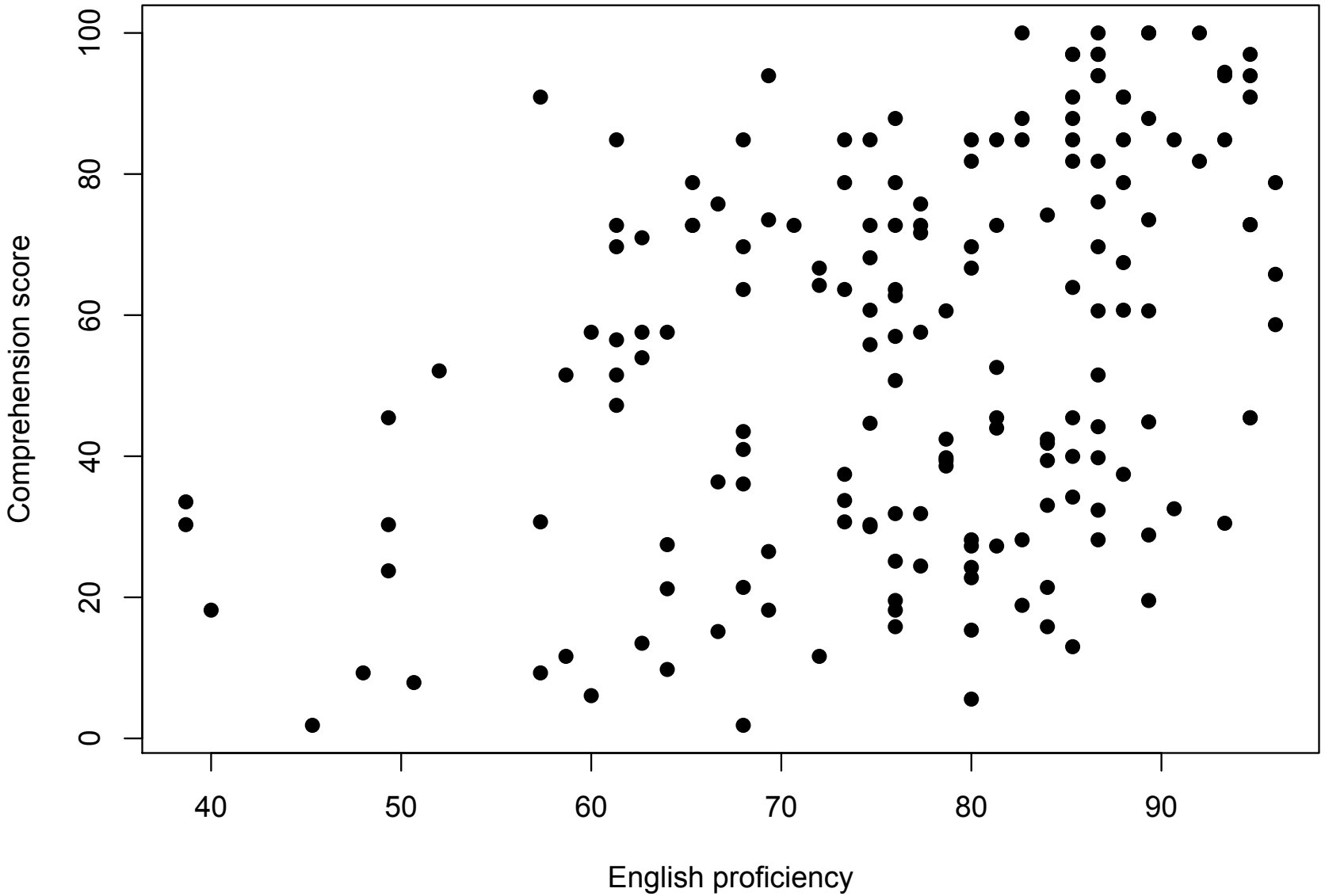
2. When the ventricles relax, the blood from the right atrium flows, through the tricuspid valve, into the right ventricle.

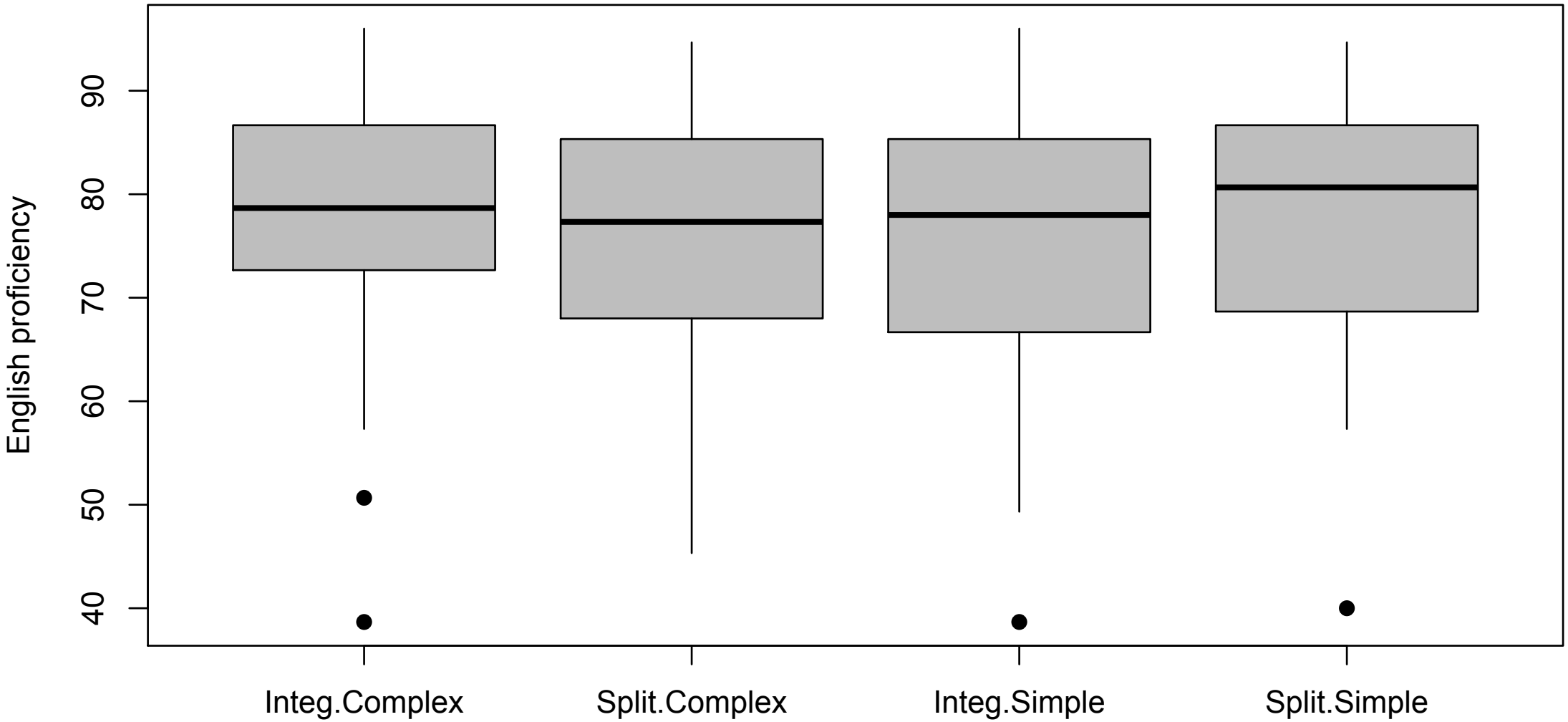
6. At the same time, blood from the left atrium flows into the left ventricle through the mitral valve.

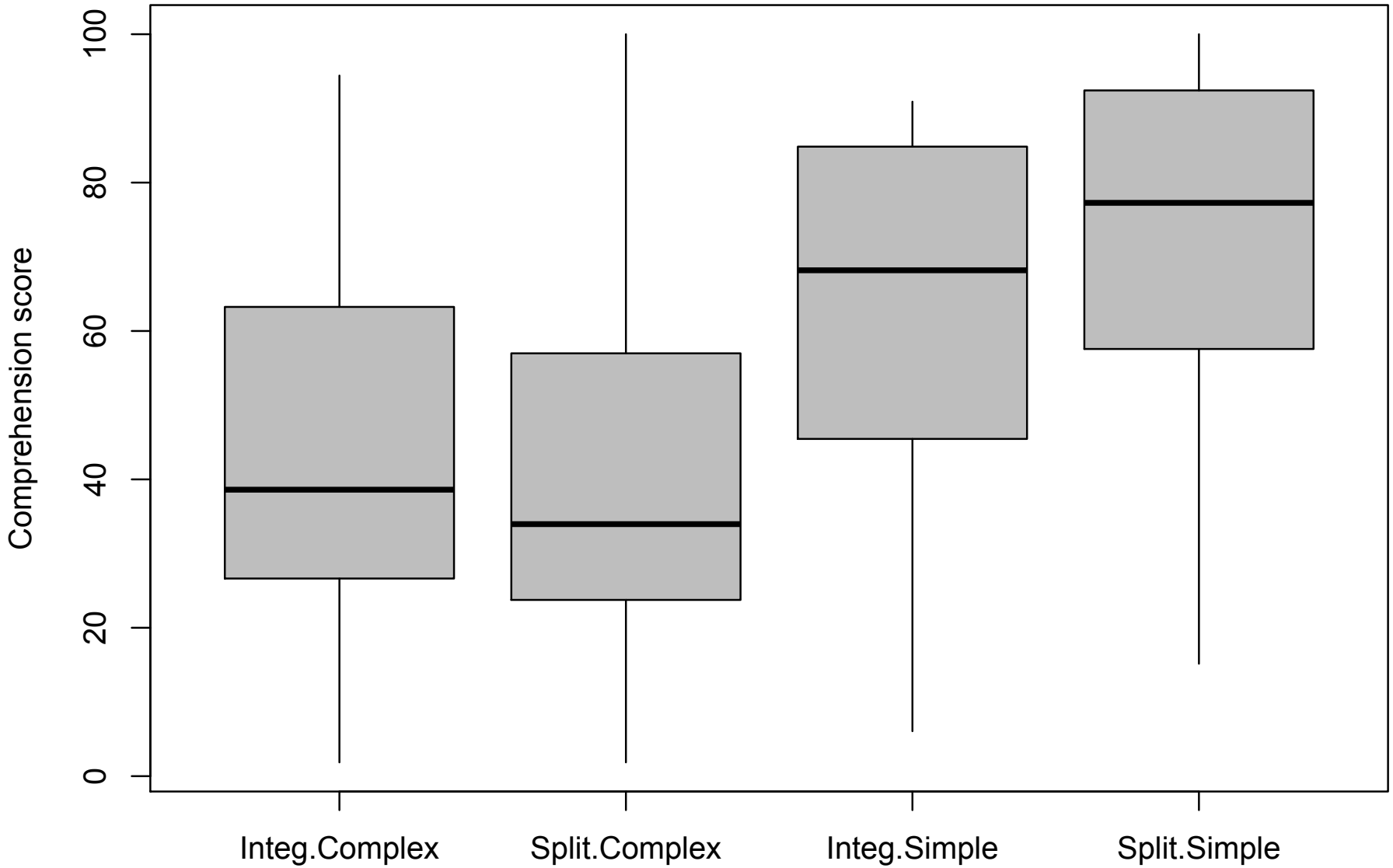


Variables

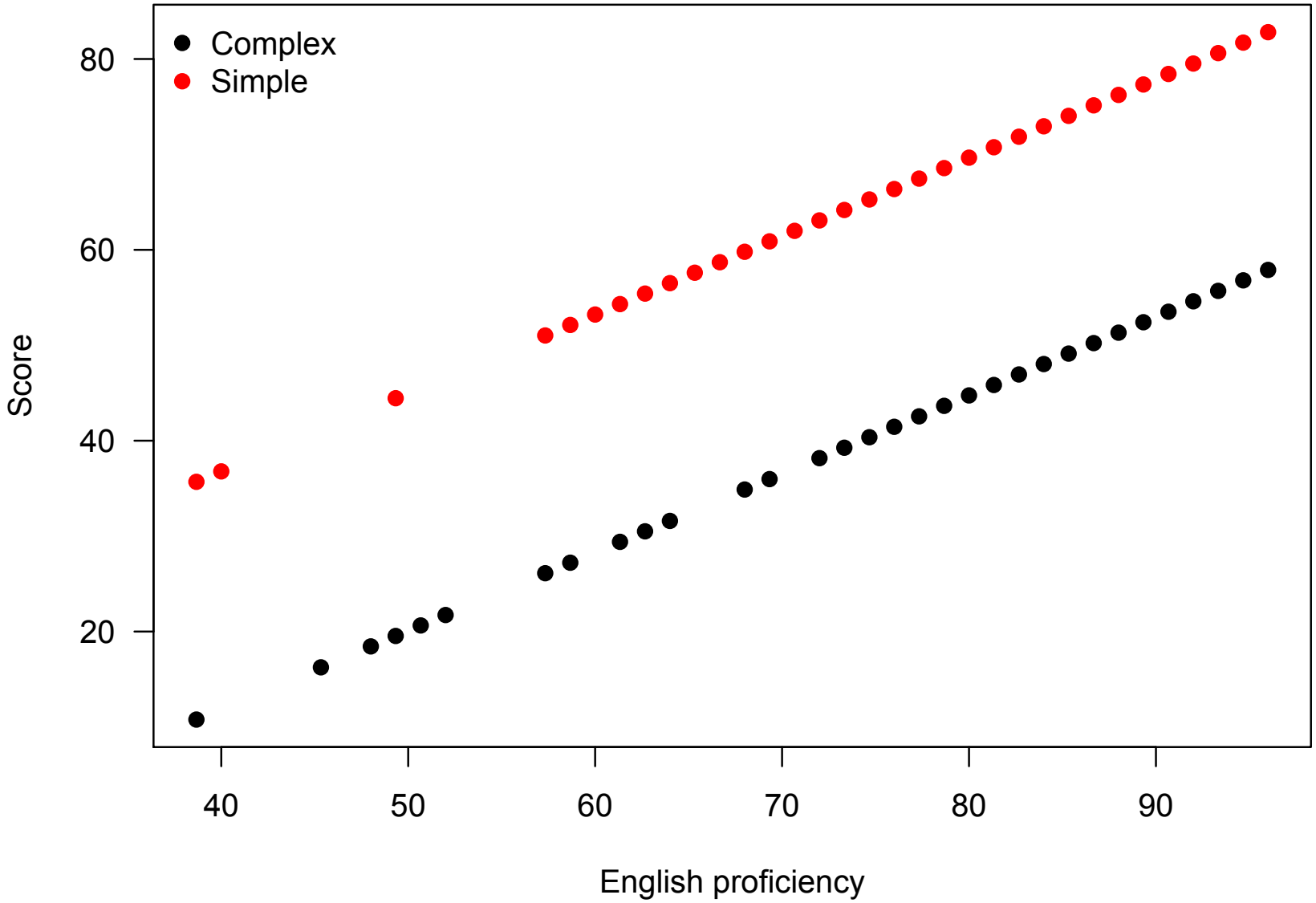
- Dependent variable
  - text comprehension (numerical)
- Independent variables
  - Layout (categorical: integrated or split)
  - difficulty (categorical: complex or simple)
  - English proficiency (numerical)


Comprehension score
0
20
40
60
80
100
40
50
60
70
80
90
English proficiency


English proficiency
40
50
60
70
80
90
Integ.Complex
Split.Complex
Integ.Simple
Split.Simple


Comprehension score
0
20
40
60
80
100
Integ.Complex
Split.Complex
Integ.Simple
Split.Simple

Results


Complex
Simple
80
60
40
20
Score
40
50
60
70
80
90
English proficiency

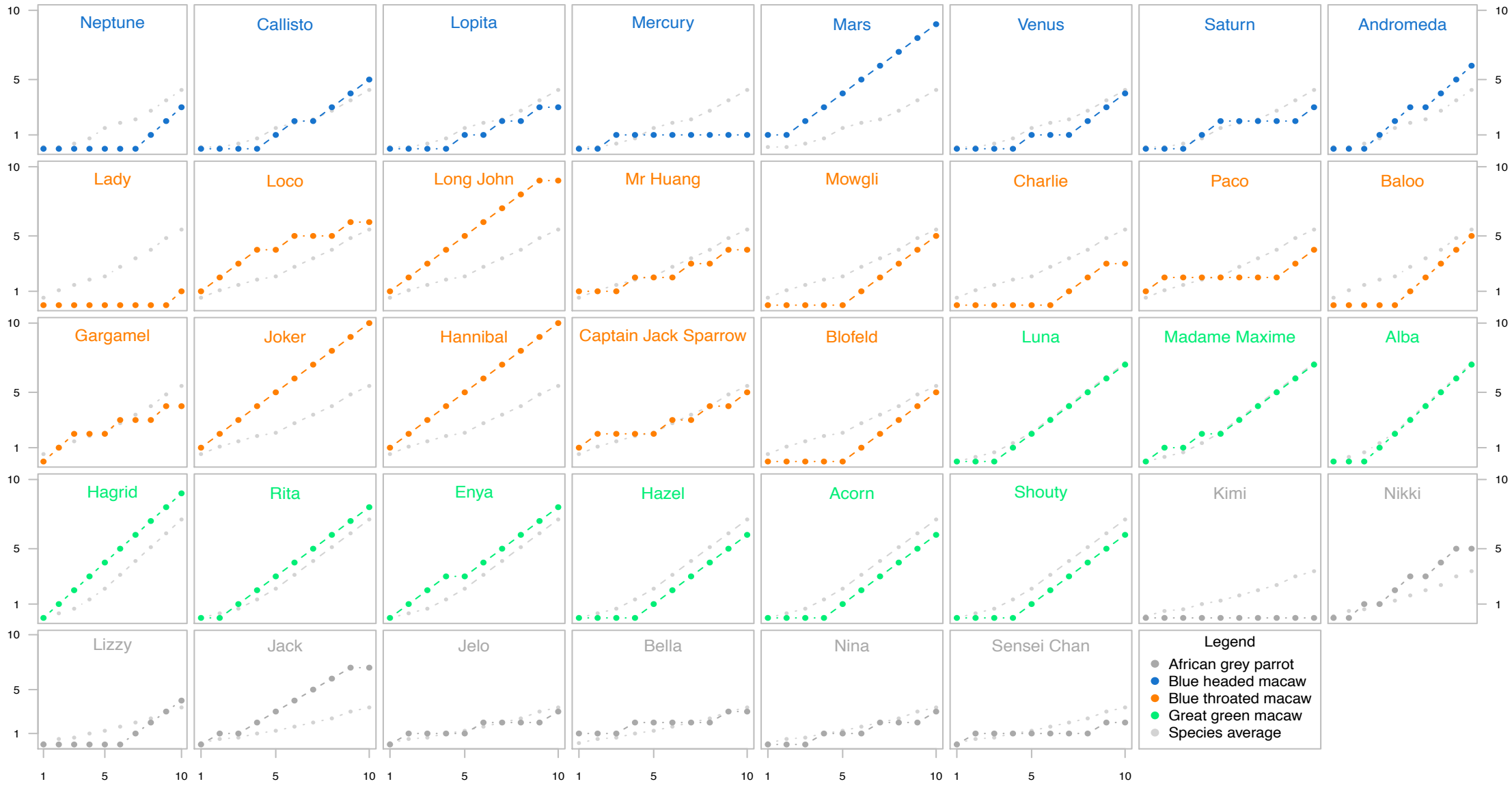
Example study 2: Parrots with Can Kabadayi

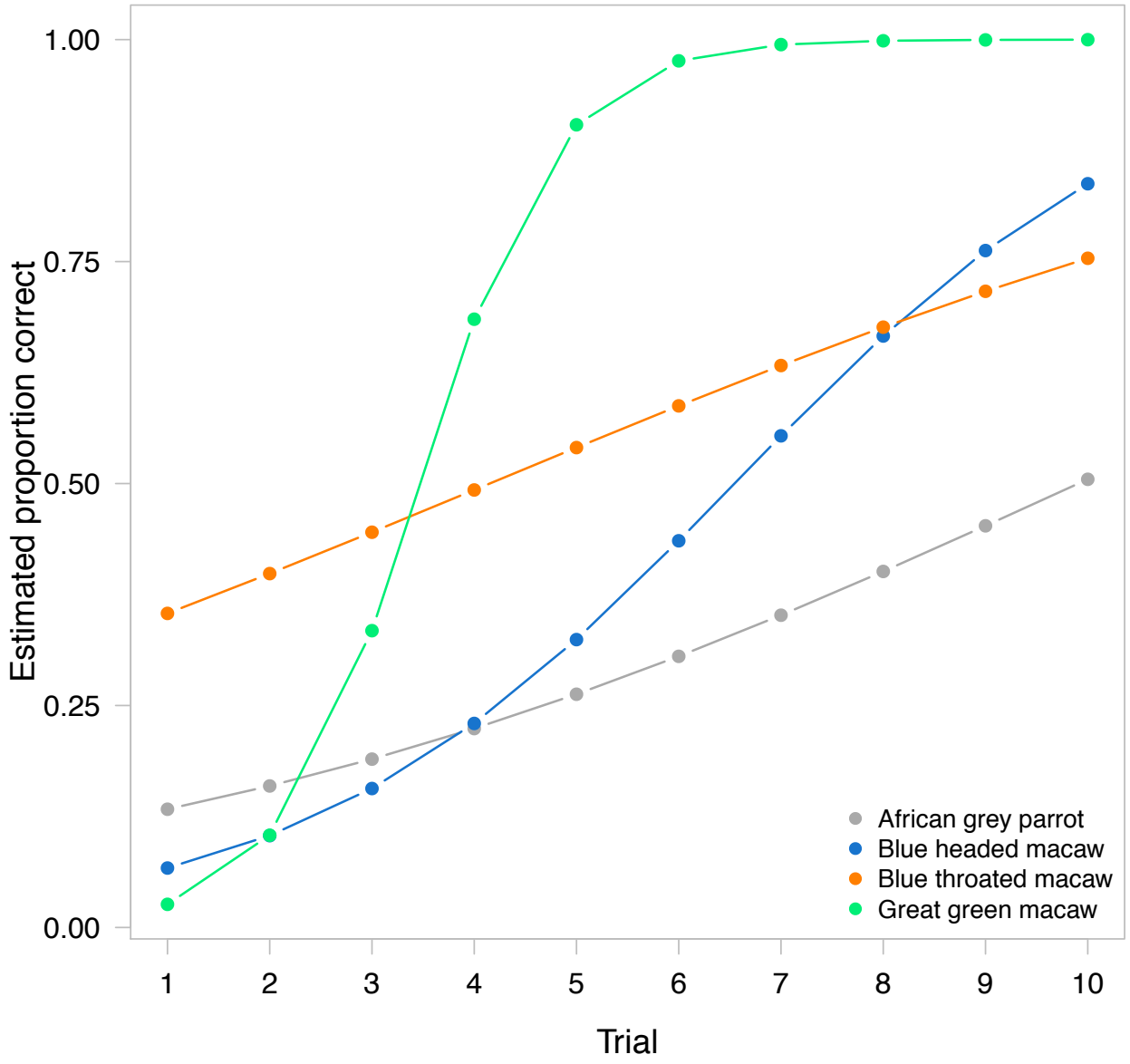
Background

- In a fairly recent study on animal cognition (Maclean et al, 2014) , several different species performed the 'cylinder task'.
- In this task, an animal sees food inside a transparent tube. The animal's spontaneous response is to reach for the food 'through' the wall of the tube, which leads to a failure. However, after several trials, some animals learn to inhibit this response and go to the sides of the tube immediately.
- A result of the study was that animals with larger brains (notably, primates) were better at the task than animals with smaller brains, and therefore it was concluded that there is a relation between brain size and the cognitive skill of inhibiting responses.
- However, some time later (Kabadayi et al. 2016), the study was replicated with corvids (crows), and it was found that they performed equally well as the primates in the earlier study, in spite of having much smaller brains.
- The question was raised what the relationship is between brain size and animal cognition.
- Additional bird species needed to be tested!

Study design

- Four different parrot species were tested on the cylinder task.
- Each bird was given ten trials to the task.
- Responses were coded as a success or a failure.
- There were different numbers of birds within each species.
- Dependent variable
  - success rate
- Independent variables
  - bird species (categorical),
  - trial number (numerical)


Neptune
Callisto
Lopita
Mercury
Mars
Venus
Saturn
Andromeda
Lady
Loco
Long John
Mr Huang
Mowgli
Charlie
Paco
Baloo
Gargamel
Joker
Hannibal
Captain Jack Sparrow
Blofeld
Luna
Madame Maxime
Alba
Hagrid
Rita
Enya
Hazel
Acorn
Shouty
Kimi
Nikki
Lizzy
Jack
Jelo
Bella
Nina
Sensei Chan
10
5
1
Legend
African grey parrot
Blue headed macaw
Blue throated macaw
Great green macaw
Species average


1.00
0.75
0.50
0.25
0.00
Estimated proportion correct
1
2
3
4
5
6
7
8
9
10
Trial
African grey parrot
Blue headed macaw
Blue throated macaw
Great green macaw

Confound: Age (years)

- African grey parrot: 1, 1, 1, 1, 1, 1, 1, 1
- Blue-throated macaw: 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 4, 4, 7
- Great green macaw: 1, 1, 1, 1, 1, 1, 2, 4, 9
- Blue-headed macaw: 1, 1, 1, 1, 1, 1, 1, 1

Example study 3: Writing intervention with Birgitta Sahlén

Background and questions

- Observational learning is a learning strategy where a pupil observes a fellow pupil performing a task and learns from this in the absence of explicit instruction.
- The method has been successfully applied to secondary school pupils, and highschool students, especially in the area of writing an essay. In this area, a student is recorded on video reading a text written by a fellow student, and makes comments on parts of the text that are unclear or unclear.
- The question is whether the method also has positive effects on hearing-impaired students.

Design of the study

- Three groups of university students participated, one control group of 10 normal-hearing university students who did not watch the video, one group of 10 normal-hearing university students who did watch the video, and one group of four hearing-impaired students who also watched the videos.
- The participants wrote texts before and after watching the videos. The quality of the texts was judged by a panel of raters, to see whether there was an improvement in the experimental groups but not in the control group.

Table 2: Results of the statistical analysis. Explanation is provided in the text.

	NI-NH in text AO		NI+NH-CT in text AO		AO-CO in group HI		change in the difference between NI-NH in text CO		change in the difference between NI+NH-CT in text CO.	
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
rating	-0.543	0.593	-0.590	0.562	-0.418	0.680	0.661	0.516	1.048	0.307
word tokens	-0.937	0.360	0.004	0.997	-1.610	0.122	0.303	0.765	1.271	0.218
VocD	0.383	0.706	**2.044**	**0.054**	0.599	0.556	-0.125	0.902	-1.078	0.293
lexical density	0.955	0.351	1.116	0.277	-0.633	0.533	0.673	0.508	-1.441	0.164
words per clause	**2.127**	**0.045**	0.549	0.589	-0.361	0.722	-0.990	0.334	-0.992	0.333
clauses per T-unit	-1.576	0.130	0.200	0.843	**-2.131**	**0.045**	1.363	0.187	1.277	0.216
pause time	0.342	0.736	1.044	0.308	0.759	0.457	-0.357	0.724	-0.830	0.416
characters deleted	0.698	0.493	-0.248	0.806	**3.375**	**0.003**	-1.230	0.232	**-1.902**	**0.071**
characters per second	-1.419	0.171	-0.215	0.832	-1.473	0.156	0.404	0.690	0.669	0.511

Example study 4: Case shift with Halldor Sigurdsson

Background and questions

- Case system in Swedish is impoverished compared to some other Germanic languages. There is case on some personal pronouns (contrasting nominative with oblique case) but not on nouns or determiners.
- There is a trend of case shift, where the oblique personal pronoun is being replaced by the nominative case (e.g., *jag ser han*, 'I see he') or the other way around (e.g., *det är mig*, 'it is me').
- A survey (sentence acceptability judgments) was sent out to native speakers to investigate which factors were driving these examples of case shift, for instance, speaker age, syntactic construction, syntactic co-ordination, etc.

Example study 4: Pantomime with Jordan Zlatev

Background and questions

- There is controversy as to how language came into existence: from gestures or from (non-verbal) vocalizations.
- One way of approaching this controversy is by asking how communicatively successful these strategies are.
- Participants were shown short video clips showing actors performing transitive events (e.g., *boy hits man*). The actors either used vocalizations or not.
- The participants were asked to identify the scene by selecting it from a response sheet with a set of 20 images.

1

2

3

4

5

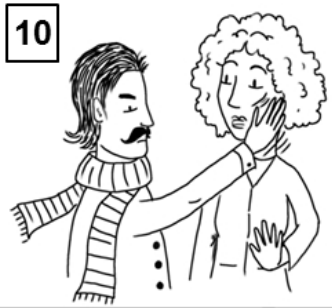
6

7

8

9

10



11

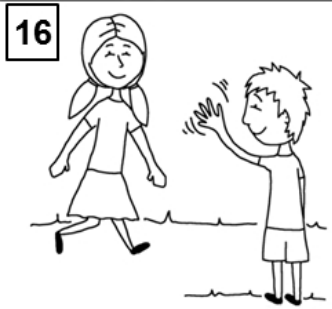
12

13

14

15

16



17

18

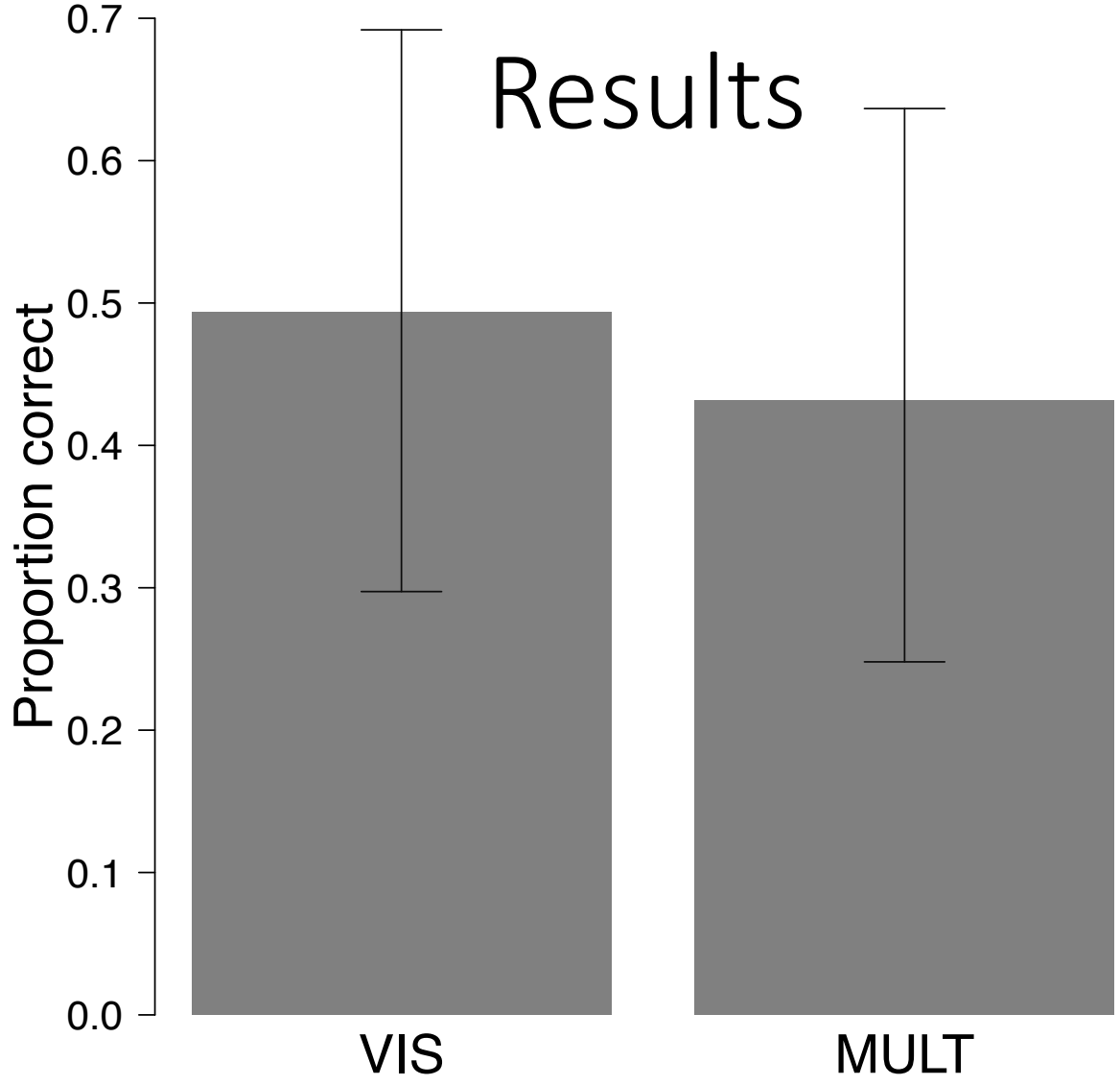
19

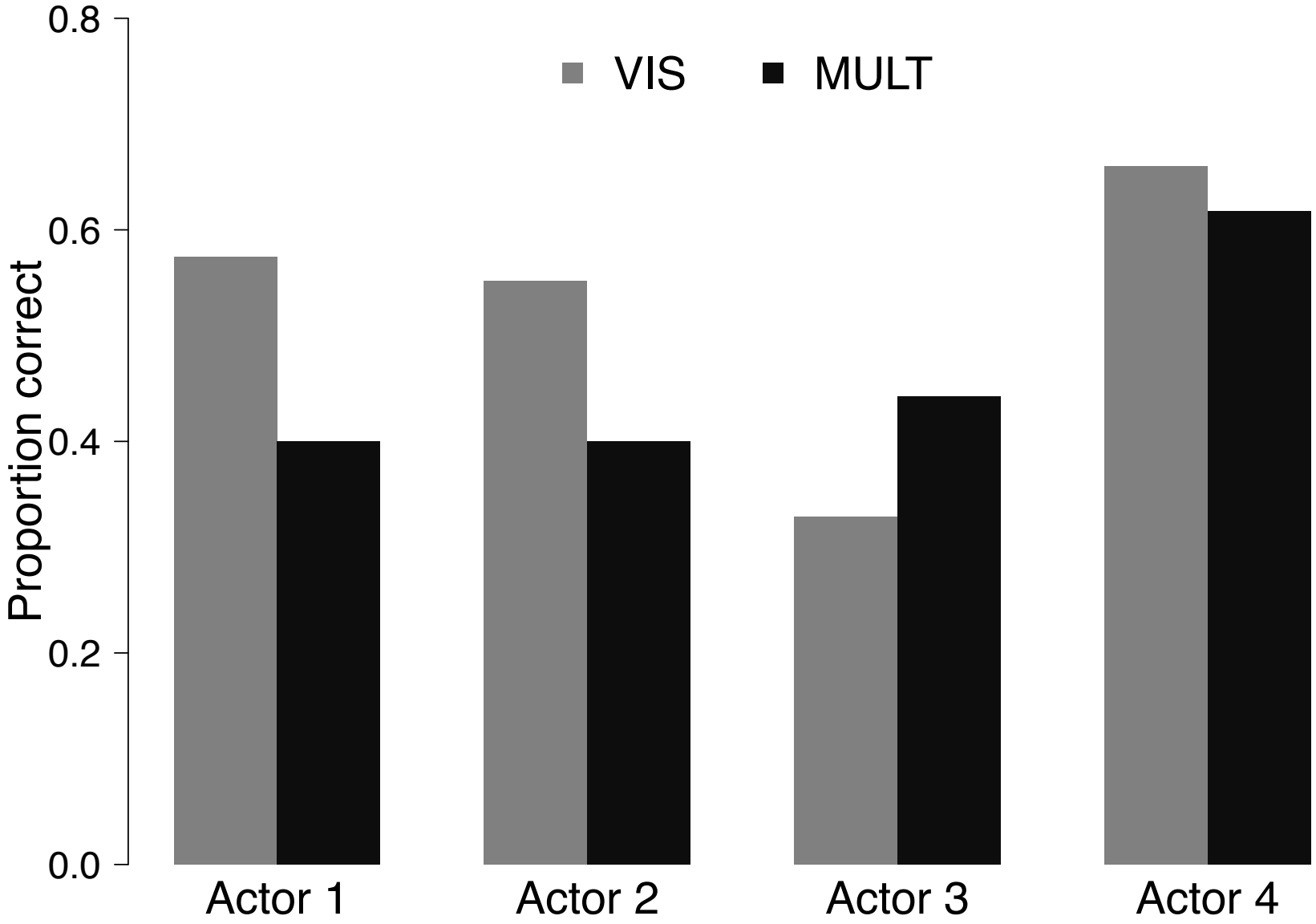
20

Study design

- The actions were performed by four different actors, in both conditions (i.e., with and without vocalizations)
- The clips were shown to an audience of 44 participants.
- There were four different versions showing the four actors performing the scenes either with or without vocalizations (counterbalanced)
- Dependent variable
  - Identification (correct or incorrect)
- Independent variables
  - condition (with or without vocalizations)
  - trial number (learning effect)

Results




0.8
0.6
0.4
0.2
0.0
Proportion correct
VIS
MULT
Actor 1
Actor 2
Actor 3
Actor 4

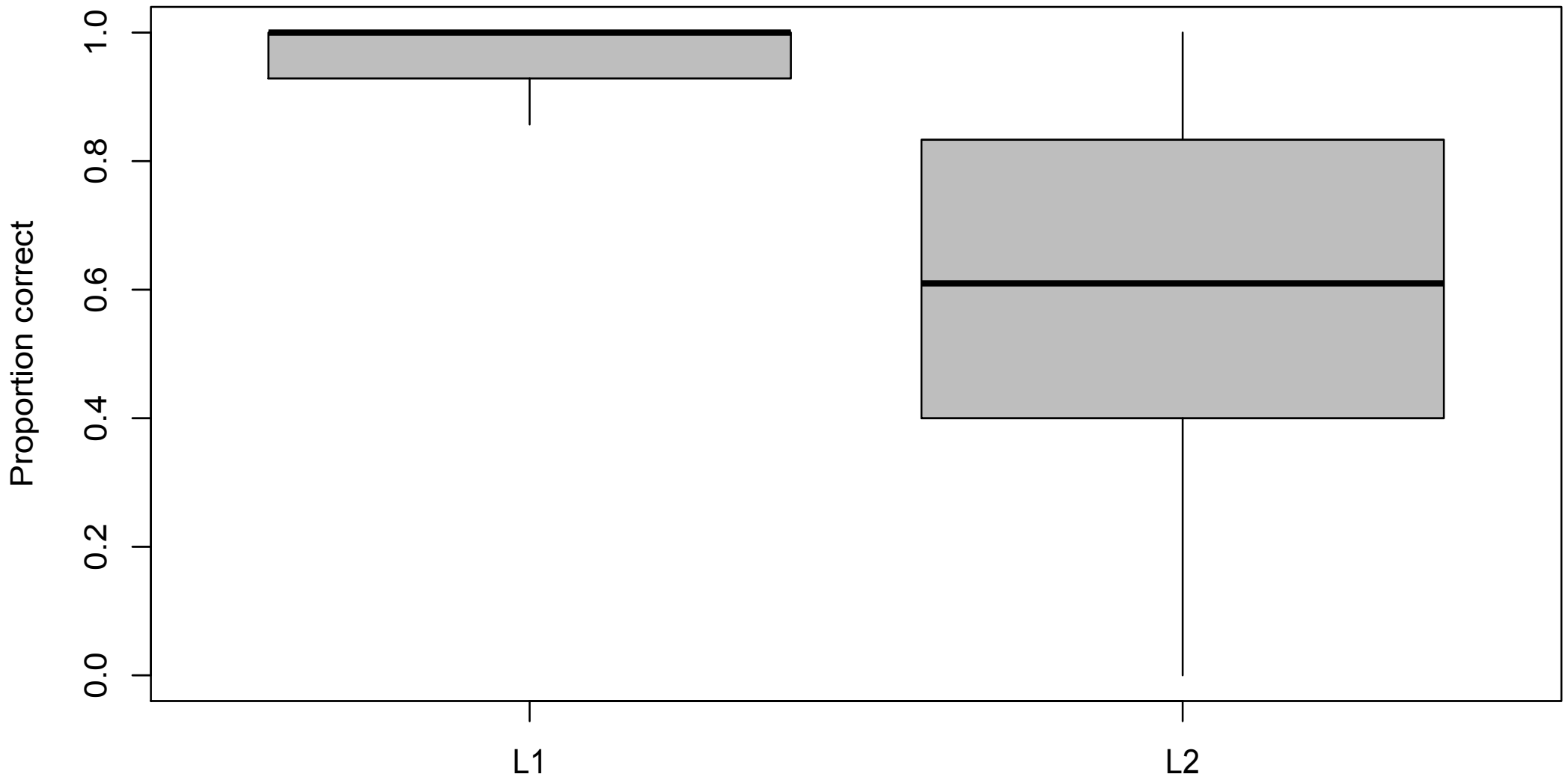
Example study 6: liaison with Malin Ågren

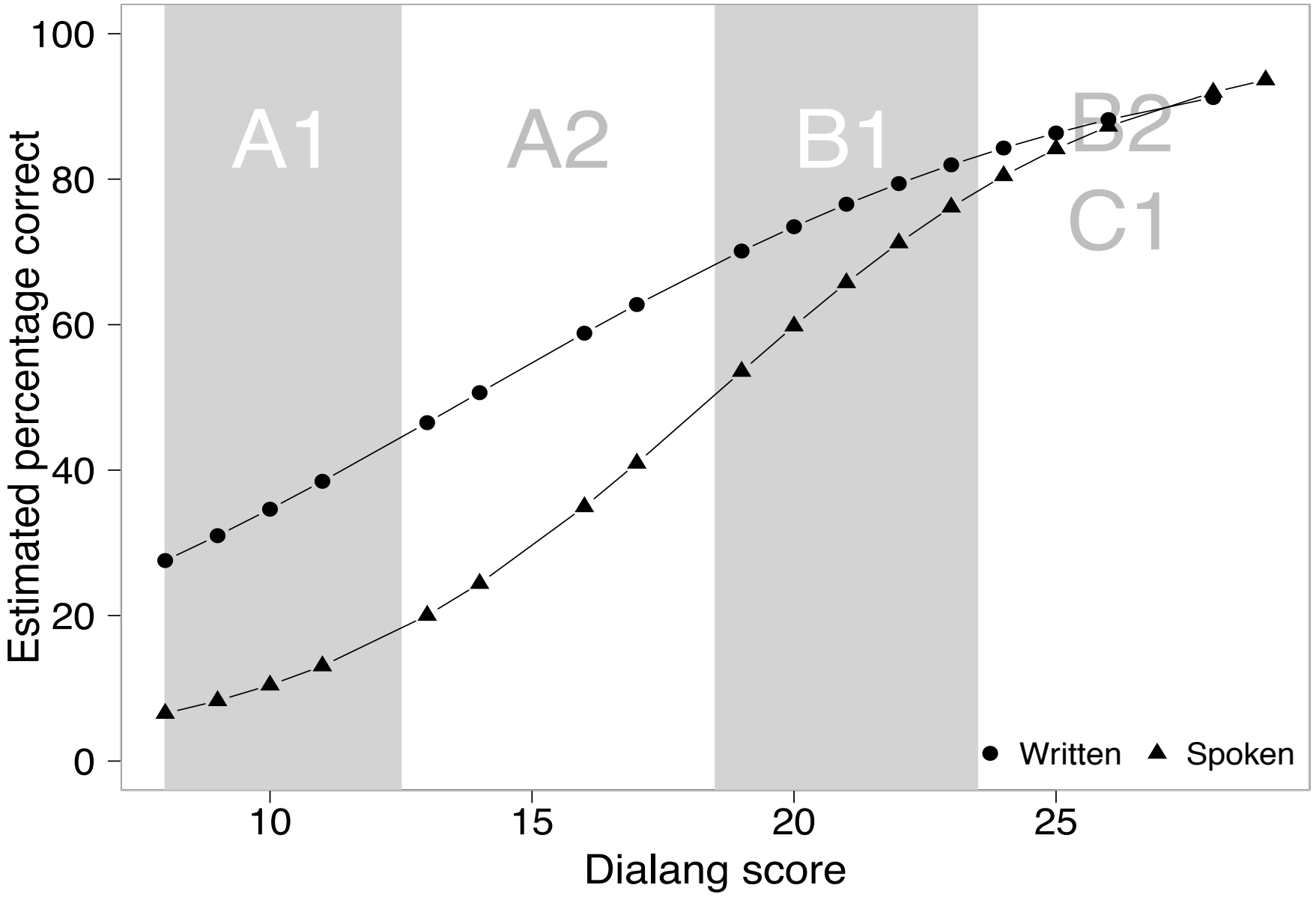
Background and questions

- Liaison is a difficult aspect of French phonology for various reasons. Sometimes liaison is obligatory, sometimes it is optional, and sometimes it does not occur at places where you expect it. Furthermore, liaison is confusing because it effectively shifts the boundary between two words to the front of the liaison consonant rather than after it.
- When do L2 learners become proficient in the production of liaison?
- Is it easier for L2 learners to reproduce an utterance with liaison when it is presented in oral mode or in written mode?
- Is liaison easier to produce on high-frequency words than on low-frequency words?

Design of the study

- A set of stimulus sentences containing liaison (plural pronoun plus verb) was created.
- The sentences were presented to a group of 44 L2 learners and 21 L1 speakers.
- The L2 learners had different levels of proficiency (as measured by a vocabulary test) and on the basis of their score on the test were classified into different learner stages.
- Half of the sentences were presented in orally, the other half written.
- The verbs had varying frequencies, based on corpus counts and participant judgments.


Proportion correct
0.0
0.2
0.4
0.6
0.8
1.0
L1
L2


A1
A2
B1
B2
C1
Written
Spoken
Estimated percentage correct
Dialang score
0
20
40
60
80
100
10
15
20
25

Conclusions, recommendations

- Formulate questions before the data collection.
- Formulate predictions for the results.
- Formulate what you will conclude if the results are confirmed, but also if the results are not confirmed.
- Make a list of potential problems that you may run into during the data analysis.
- Be critical as to how variables are being measured in the study.
- Decide on the analysis that you want to use.

Thank you!