



# **BULATS – A TOOL FOR MEASURING AND COMPARING PROGRESS IN LANGUAGE LEARNING?**

**KATHRIN HAHN, GOETHE-INSTITUT (ZENTRALE,  
BEREICH 44)**

**DR. WINFRIED KOCH, BDS KOCH**

**GOETHE  
INSTITUT**

Sprache. Kultur. Deutschland.

# THE PROJECT TABLET-CLASSROOM

- Pilotproject for the implementation of tablets (iPads) in the forgein language classroom,
- Multimedial equipment for the Goethe-Institutes London, Dublin and Amsterdam,
- Preparation of special sequences and materials using tablets in class,
- Special training für the „iPad-Teachers“ at the test Goethe-Insitutes,
- Monitoring throughout the whole project: BULATS test for listening comprehension as one part of the monitoring.

# **HOW DO TABLETS CHANGE LANGUAGE TEACHING AND LEARNING?**



Mobility in the classroom



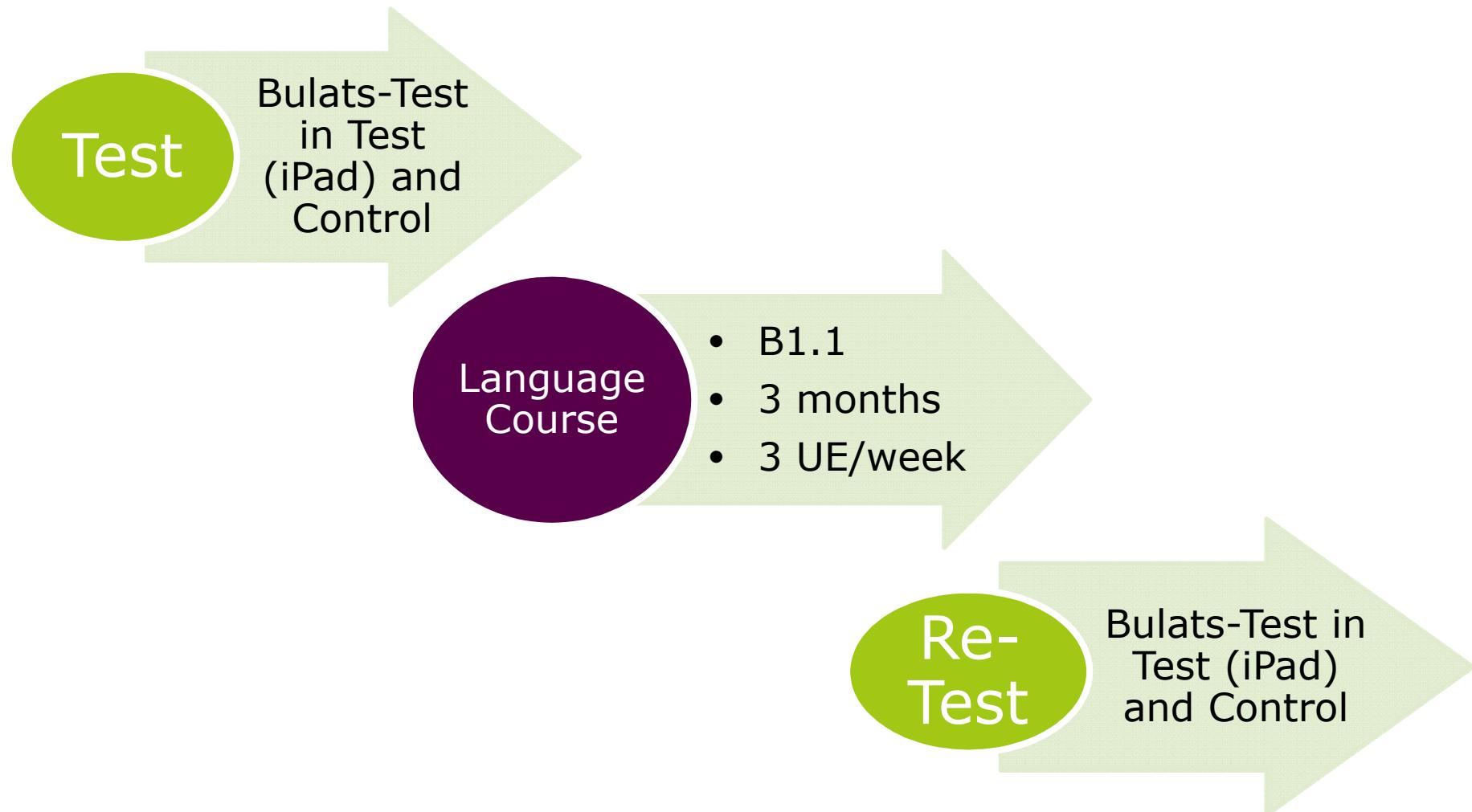




**Listening comprehension**



# RESEARCH DESIGN



# NUMBER OF LEARNERS BY GOETHE-INSTITUTE

Goethe-Institute	Class		
	Test	Control	All
Amsterdam	12	11	23
Dublin	6	4	10
London	9	6	15
All	27	21	48

# BULATS – FACTS AND FIGURES

Seite 11

16.04.2014

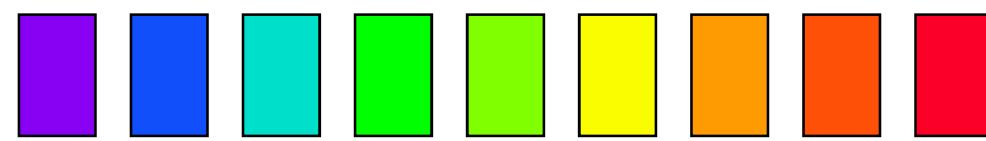
- ✓ since 1997, the Goethe-Institut is a BULATS-Partner from Cambridge ESOL, der Universidad Salamanca, der Alliance française
- ✓ BULATS ist orientated on the Common European Framework of Reference for Languages
- ✓ available for German, English, Spanish and French
- ✓ since 2008 adaptive online-version of BULATS:
  - all participants start with an introductory questions at the same level. The program chooses each test-item on the basis of the previous answer. The level of difficulty is adapted to the learner's level of language.

# BULATS – FORMAT FOR TESTING

Seite 12

16.04.2014

✓ Scaled (A1-C2)

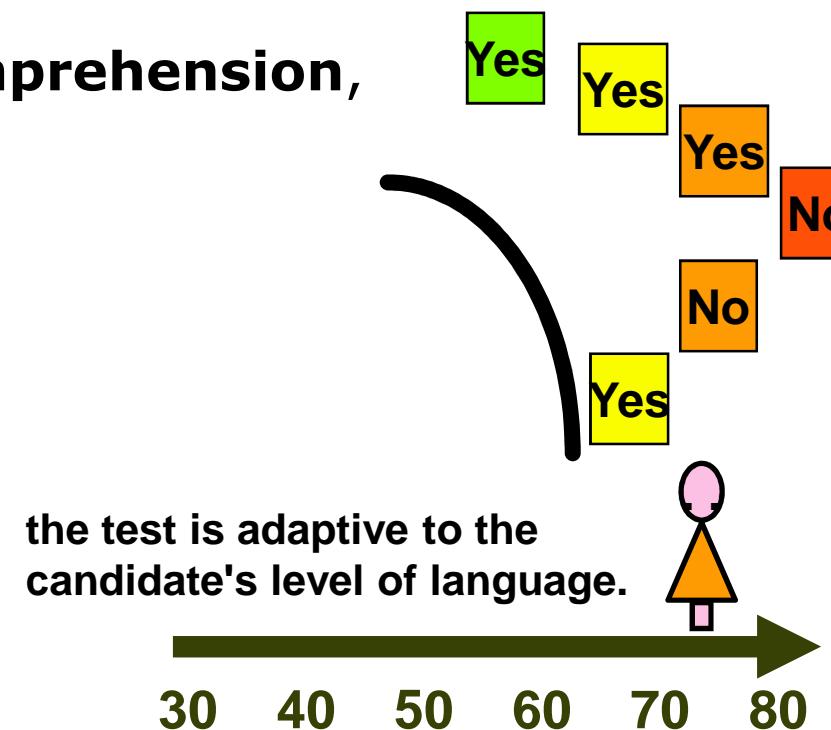


✓ Adaptive

✓ Online

✓ Reading, **Listening Comprehension**,

Vocabulary und Grammar



# BULATS – POINTS PER LEVEL

Seite 13  
Goethe-Institut für Thema  
16.04.2014

Level	0-A1*	A1	A2	B1	B2	C1	C2
Points	0-9	10-19	20-39	40-59	60-74	75-89	90-100

- ✓ the points scored in the test are represented in a scale from 0-100
- ✓ the points in the scale represent the different levels of the Common European Framework of Reference for Languages as shown in the table

\* 0-A1 includes those learners who are on their way to level A1

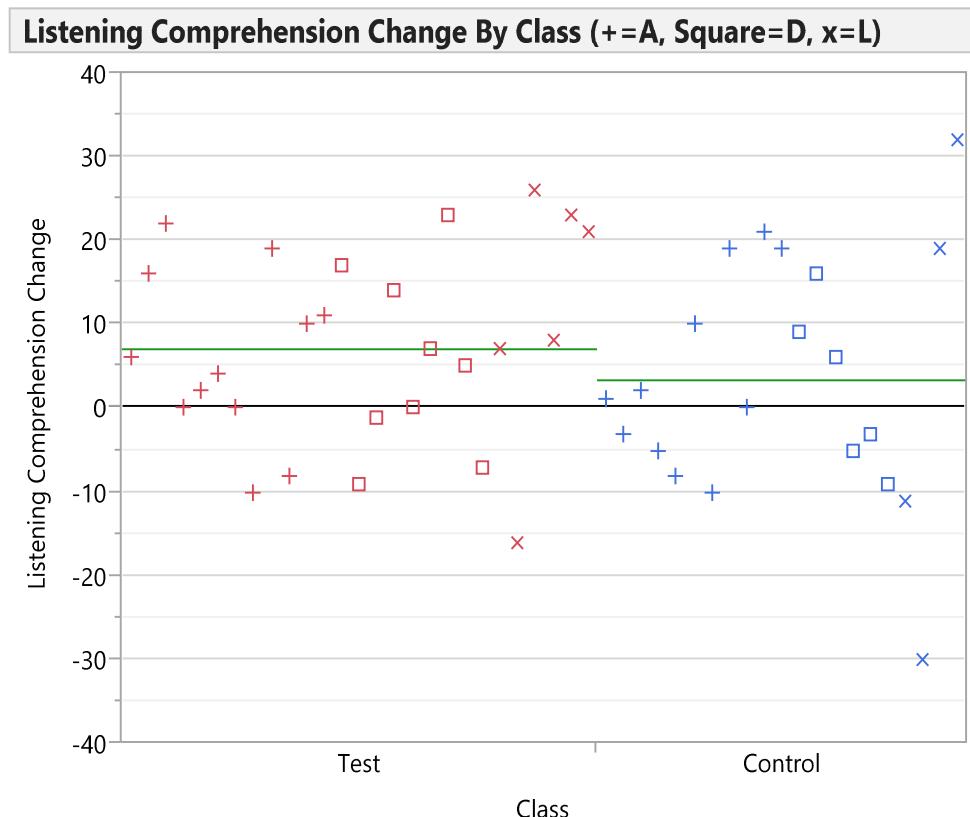
# OBJECTIVES OF STATISTICAL ANALYSIS

Seite 14

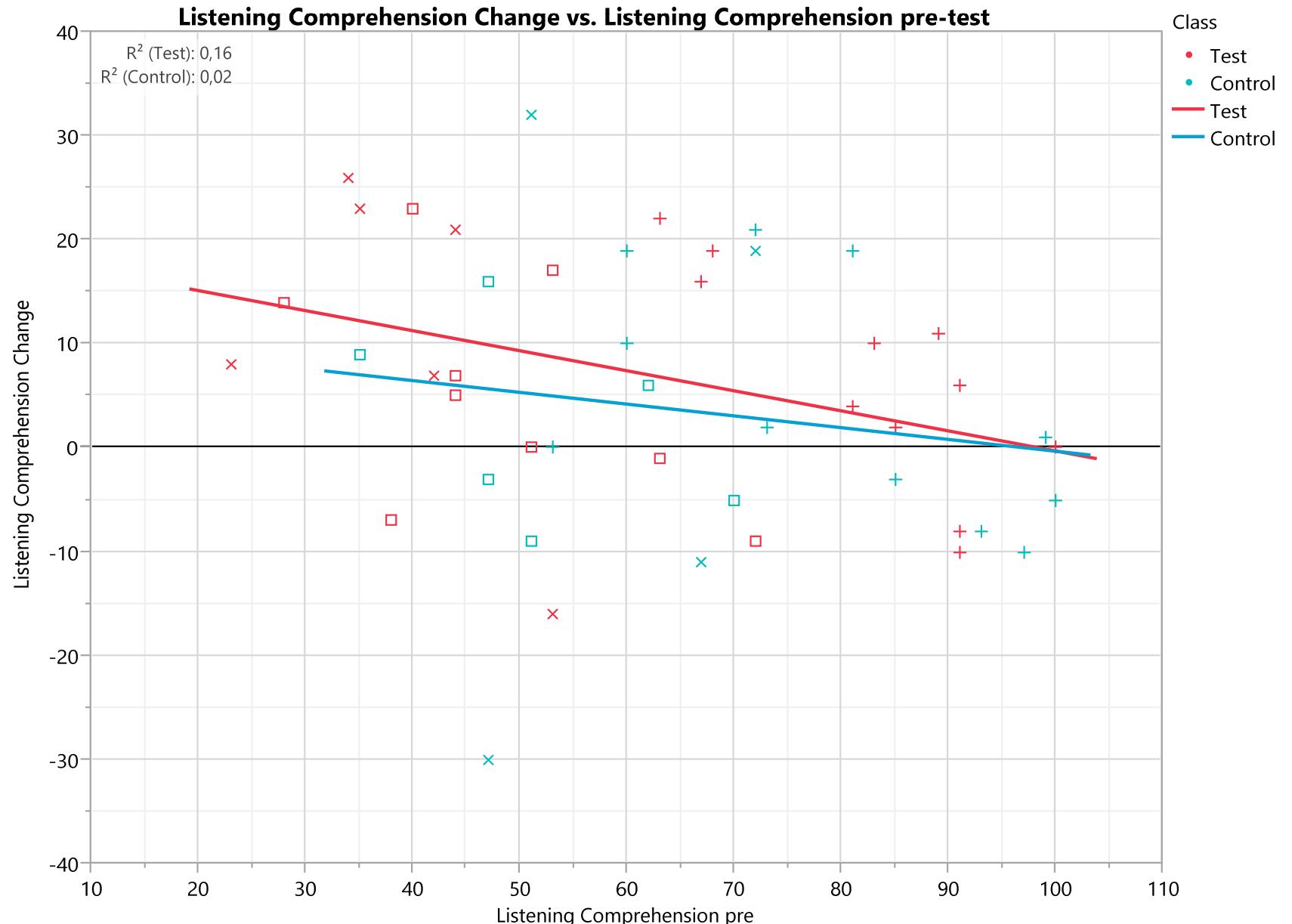
- Identifying Change/Progress in Listening Comprehension (LC)
- Null Hypothesis:  
No difference in change/progress in LC between Test (iPad) and Control Group
- Influence of the LC-results in the first test and of the Goethe-Institute tested
- Learn our lesson for future studies

# STATISTICAL TESTING OF THE NULL-HYPOTHESIS

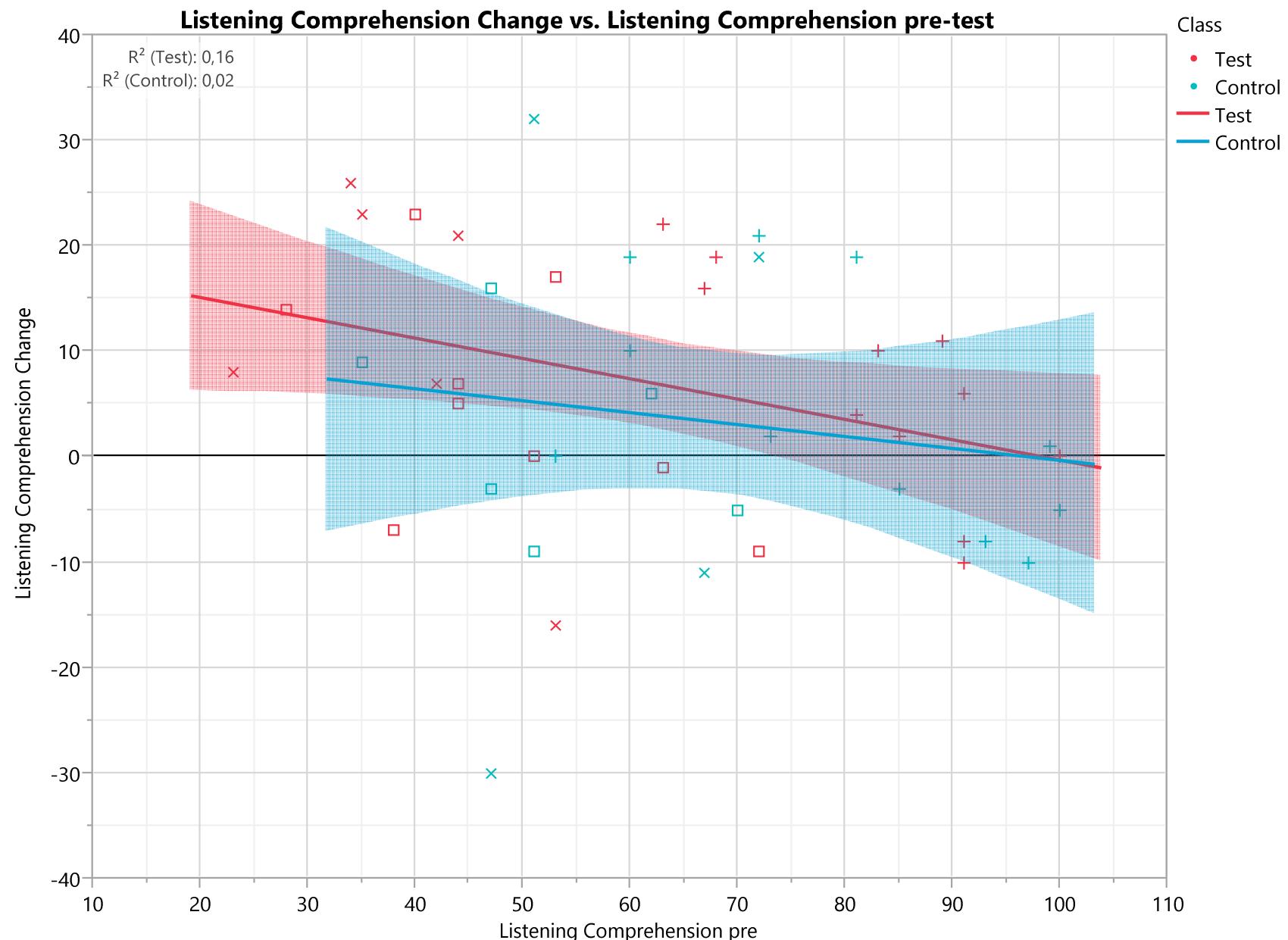
- Mean Difference = 3.7  
95%-C.I. [-3.8, 11.2]  
n=48 (27 Test, 21 Control)
  - Two-sided p values:
    - t-test p = 0.326
    - Wilcoxon test p = 0.298
- => Null-Hypothesis not rejected



## Dependency of LC Change from pre-test level



## Dependency of LC Change from pre-test level

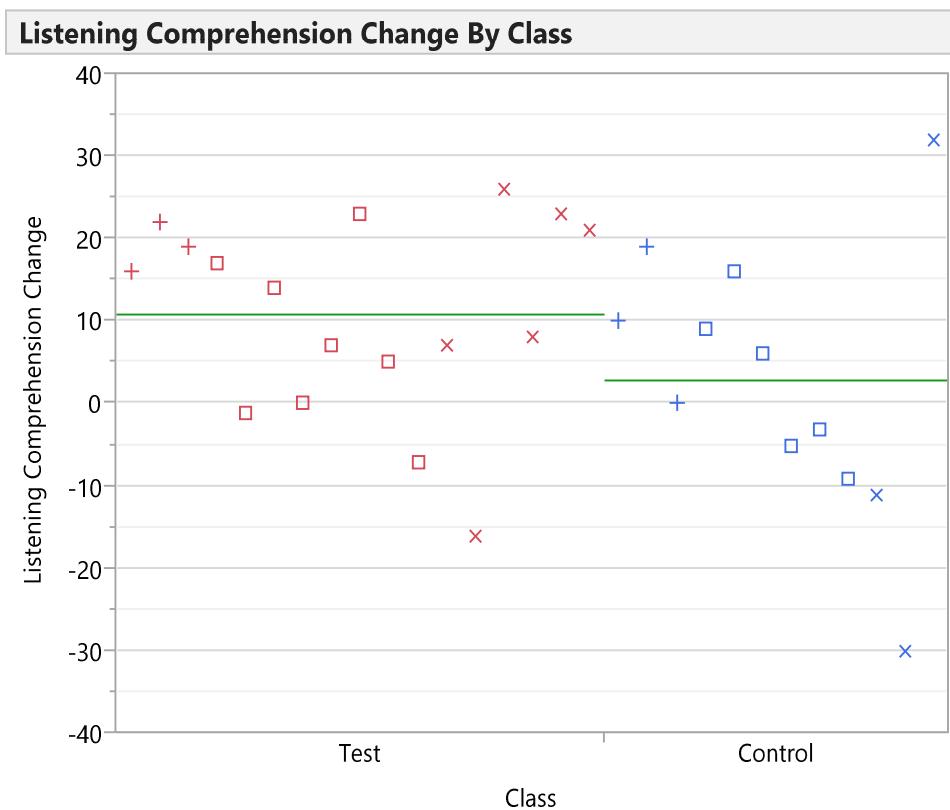


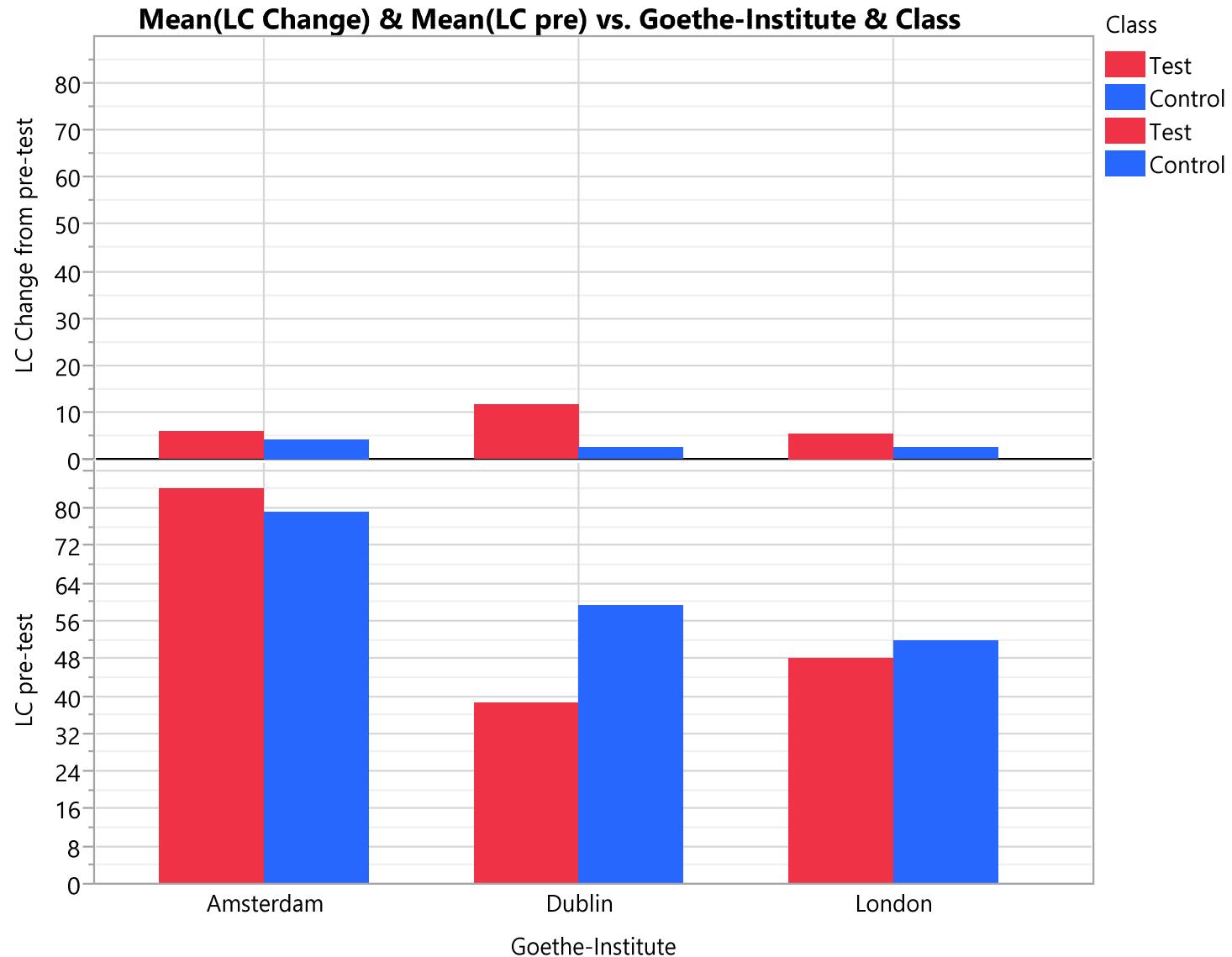
## DEPENDENCY OF CHANGE OF LC FROM PRE-TEST LEVEL

- Progress in LC decreases with increasing pre-test level,
- Above pre-test level of 80: not enough lee-way for improvement
- Up to a LC pre-test level of up to 70 the 95%-confidence bands of the linear fit of the Test Group are completely above 0 → enough leeway for differences to develop

# TESTING OF THE NULL-HYPOTHESIS FOR LEARNERS WITH PRE-TEST $\leq$ 70

- Mean Difference = 8.0  
95%-C.I. [-2.7, 18.7]  
n=29 (17 Test, 12 Control)
- t-test p = 0.137  
Wilcoxon test p = 0.144
- Relative Effect Size = 0.578
- ⇒ Null-Hypothesis not rejected but p<0.2,
- ⇒ Mean difference of 8.0 relevant,
- ⇒ Signal of a relevant effect detected



**Listening Comprehension mean Changes and mean pre-test Values by Institute**

## **LISTENING COMPREHENSION MEAN CHANGES AND MEAN PRE-TEST VALUES BY INSTITUTE**

- Baseline LC results are well comparable in Amsterdam (n=23) and London (n=15) less well in Dublin (n=10),
- At all three Goethe-Institutes the mean increase of LC is slightly higher in the Test Group compared to the Control Group,
- Amsterdam shows a good mean increase despite a high LC mean baseline level.

# SUMMARY AND CONCLUSIONS

Seite 22

- BULATS is a suitable test to study factors that may influence learning of German,
- Significant effects of pre-test LC ( $p=0.011$ ) and Institute ( $p=0.061$ ) could be discovered,
- Use of IPad had a positive effect ( $p<0.2$  on the subset of pre-LC  $\leq 70$ ),
- Include subjects up to a BULATS pre-test level of 70 or at most 80 into future trials,
- To discover a relative effect size of 0.5 with 80% power, 64 learners per group would be needed.

**THANK YOU FOR YOUR  
KIND ATTENTION.**

**KATHRIN HAHN:**

**[kathrin.hahn@goethe.de](mailto:kathrin.hahn@goethe.de)**

**DR. WINFRIED KOCH:**  
**[winfried.koch@bdskoch.de](mailto:winfried.koch@bdskoch.de)**