Can a language aptitude test predict language outcomes?

EALTA 2017

Elina Stordell

elina.stordell@testipiste.eu www.testipiste.eu

Testipiste

New newcomers in Finland

Registrations of EU citizens and first residence permits issued to third-country nationals

2014	
EU	33 %
Russia	10 %
India	6 %
China	5 %

2016	
EU	25 %
Iraq	9 %
Russia	9 %
Afghanistan	5 %

Finnish Immigration Service http://www.migri.fi/about_us/statistics

Initial testing at Testipiste

Integration course in Finnish for jobseekers

Several different kinds of study paths available:

- Fast
- Medium
- Slow
- Literacy

Initial testing:

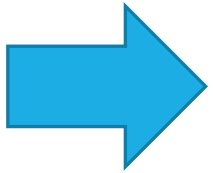
Interview

Study skills

- Technical reading
- Dictation
- **Structural reasoning**
- Math

Finnish skills

- Speaking
- Writing
- Reading comprehension
- Listening comprehension



Can the test of structural reasoning help to choose the right study path?

Can the test predict language outcomes?

Test of Structural Reasoning

Idea comes from Huis van het Nederlands, Bruxelles: COVAAR II

Finnish adaptations since 2008

Does not require any Finnish skill

Three tasks (30 items) measuring

1. **the speed and accuracy of recognizing a given word form**

Test of Structural Reasoning

Idea comes from Huis van het Nederlands, Bruxelles: COVAAR II

Finnish adaptations since 2008

Does not require any Finnish skill

Three tasks (30 items) measuring

1. **the speed and accuracy of recognizing a given word form**

Test of Structural Reasoning

Three tasks (30 items) measuring

1. the speed and accuracy of recognizing a given word form
2. **the ability to analyze structural similarities and differences**

Test of Structural Reasoning

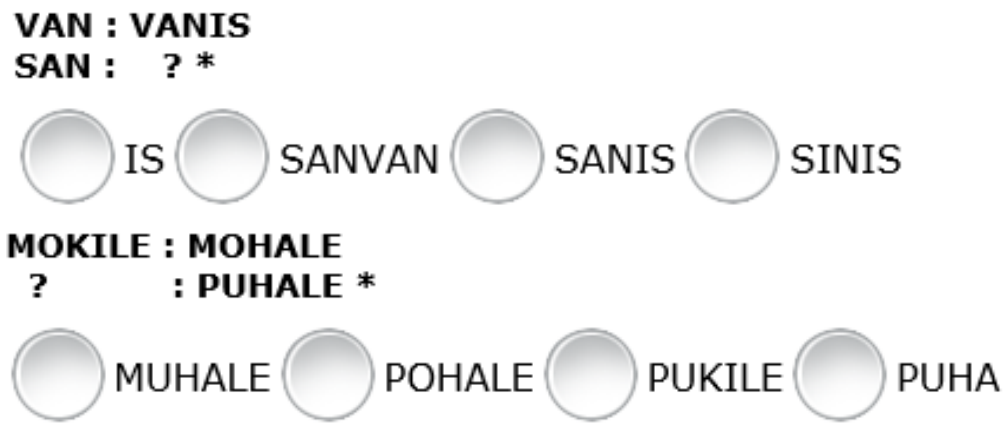
Three tasks (30 items) measuring

1. the speed and accuracy of recognizing a given word form
2. **the ability to analyze structural similarities and differences**

Test of Structural Reasoning

Three tasks (30 items) measuring

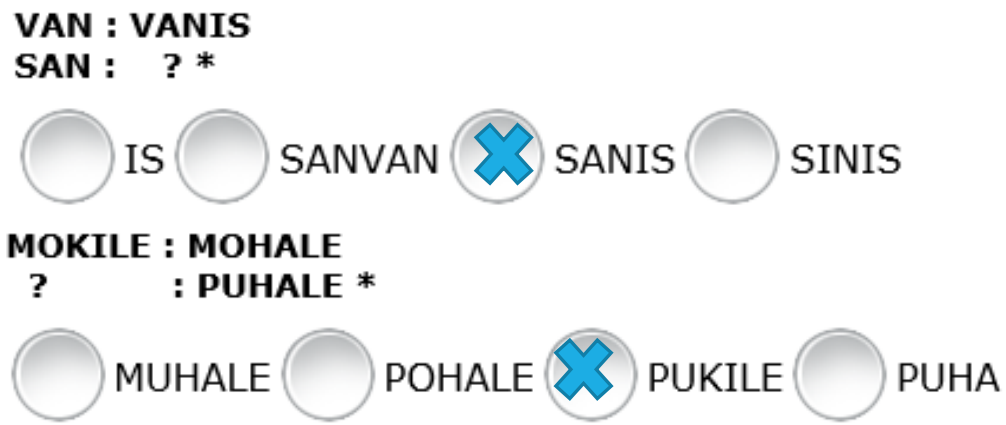
1. the speed and accuracy of recognizing a given word form
2. the ability to analyze structural similarities and differences
3. **the ability to use analogies**



Test of Structural Reasoning

Three tasks (30 items) measuring

1. the speed and accuracy of recognizing a given word form
2. the ability to analyze structural similarities and differences
3. **the ability to use analogies**



Test of Structural Reasoning

Three tasks (30 items) measuring

1. the speed and accuracy of recognizing a given word form
2. the ability to analyze structural similarities and differences
3. the ability to use analogies
4. **the ability to understand instructions and tasks**

Test of Structural Reasoning

Explained face-to-face in the interview

Examples per task

Instructions in more than 20 languages

Advice also available in the test situation

Pre-tested in different kinds of studypaths

Several comparable versions

~ 20.000 test takers since 2010

Test of Structural Reasoning: an example from the analyses

One-parameter Rasch model, Winsteps 3.80.0

N = 1880

SUMMARY OF 1880 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	21.2	30.0	1.63	.77				
S.D.	8.1	.0	2.29	.49				
MAX.	30.0	30.0	5.14	1.84				
MIN.	.0	30.0	-5.03	.41				

REAL RMSE .93 TRUE SD 2.09 SEPARATION 2.23 PERSON RELIABILITY .83
MODEL RMSE .91 TRUE SD 2.09 SEPARATION 2.29 PERSON RELIABILITY .84
S.E. OF PERSON MEAN = .05

PERSON RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .94

SUMMARY OF 30 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	1328.2	1880.0	.00	.07	.97	-1.2	1.27	1.3
S.D.	234.8	.0	1.05	.01	.15	4.4	.55	4.3
MAX.	1651.0	1880.0	2.06	.09	1.33	8.8	2.45	9.9
MIN.	846.0	1880.0	-1.69	.06	.77	-8.2	.57	-5.4

REAL RMSE .07 TRUE SD 1.05 SEPARATION 14.50 ITEM RELIABILITY 1.00
MODEL RMSE .07 TRUE SD 1.05 SEPARATION 14.89 ITEM RELIABILITY 1.00
S.E. OF ITEM MEAN = .20

ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00
47970 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 38845.44 with 46342 d.f. p=1.0000
Global Root-Mean-Square Residual (excluding extreme scores): .3541
Capped Binomial Deviance = .1502 for 56400.0 dichotomous observations
UMEAN=.0000 USCALE=1.0000

Test of Structural Reasoning: an example from the analyses

Items in the first task?

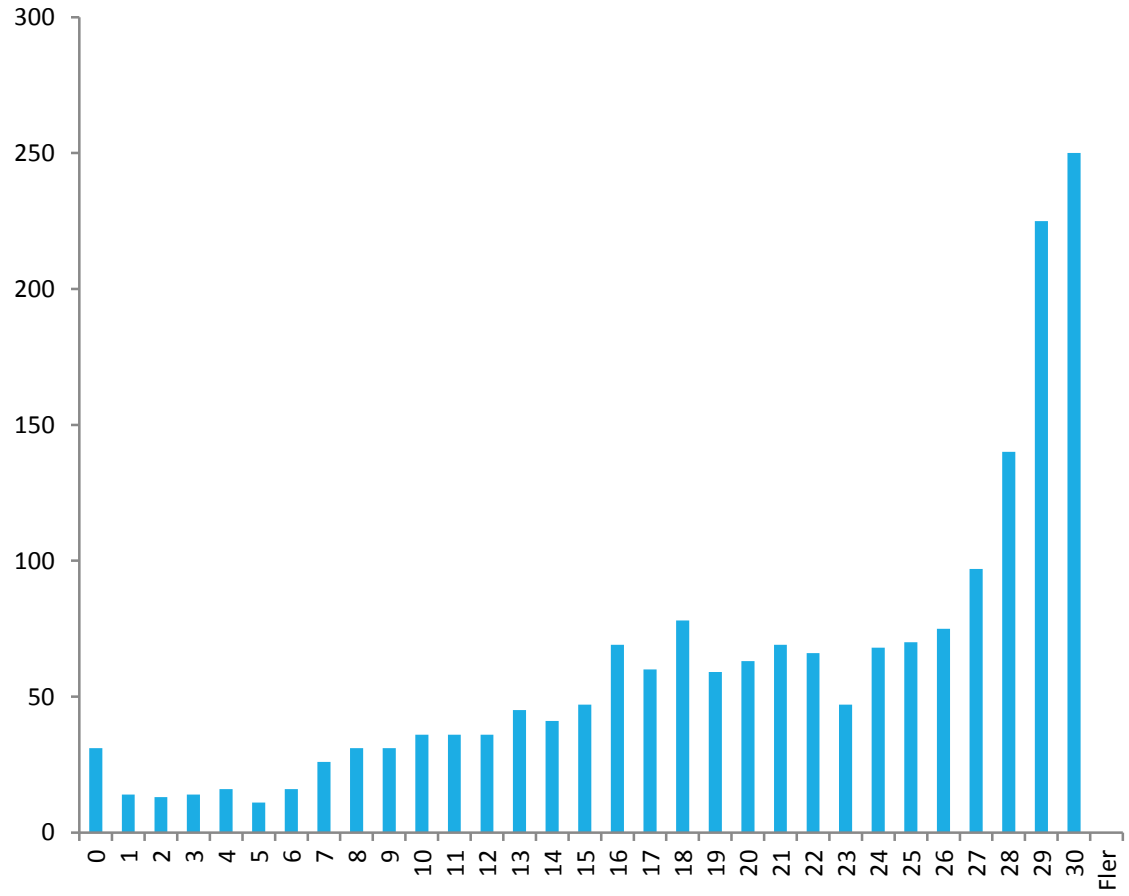
Different construct?

Time limit?

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	ESTIM DISCR
1	1415	1880	-.31	.07	1.27	7.2	1.84	7.1	.55
2	1532	1880	-.91	.08	1.10	2.3	1.87	5.4	.81
3	1580	1880	-1.19	.08	1.14	2.9	2.21	6.1	.75
4	1651	1880	-1.69	.09	.96	-.8	1.48	2.2	1.00
5	1535	1880	-.93	.08	1.17	4.0	2.12	6.6	.70
6	1582	1880	-1.21	.08	.99	-.2	1.49	2.9	.96
7	1492	1880	-.69	.07	1.20	5.0	2.30	8.3	.63
8	1471	1880	-.58	.07	1.24	6.2	2.45	9.5	.56
9	1418	1880	-.32	.07	1.33	8.8	2.30	9.9	.42
10	1523	1880	-.86	.07	1.12	2.9	1.62	4.2	.80
11	1615	1880	-1.42	.08	.85	-3.1	.57	-2.9	1.16
12	1511	1880	-.79	.07	.88	-3.2	.81	-1.6	1.14
13	1523	1880	-.86	.07	.93	-1.9	1.09	.7	1.06
14	1450	1880	-.48	.07	.92	-2.2	.89	-1.0	1.08
15	1481	1880	-.63	.07	.92	-2.3	.77	-2.2	1.11
16	1521	1880	-.85	.07	.82	-4.8	.88	-.9	1.18
17	1403	1880	-.25	.07	.94	-1.9	1.30	3.0	1.02
18	1324	1880	.10	.07	.99	-.4	1.13	1.6	.98
19	1223	1880	.53	.06	.93	-2.3	.86	-2.1	1.11
20	1181	1880	.70	.06	.92	-2.7	.91	-1.4	1.12
21	1195	1880	.64	.06	1.02	.6	1.23	3.3	.92
22	1146	1880	.84	.06	.82	-6.4	.78	-3.7	1.27
23	1198	1880	.63	.06	.80	-7.1	.69	-5.4	1.32
24	1136	1880	.88	.06	.82	-6.2	.73	-4.7	1.28
25	1046	1880	1.24	.06	.90	-3.4	.92	-1.3	1.14
26	1070	1880	1.15	.06	.77	-8.2	.72	-4.8	1.35
27	940	1880	1.67	.06	.77	-7.6	.83	-2.5	1.30
28	913	1880	1.78	.06	.81	-6.0	1.00	.0	1.23
29	924	1880	1.74	.06	.94	-2.0	1.08	1.1	1.05
30	846	1880	2.06	.07	.89	-3.2	1.22	2.6	1.09
MEAN	1328.2	1880.0	.00	.07	.97	-1.2	1.27	1.3	
S.D.	234.8	.0	1.05	.01	.15	4.4	.55	4.3	

Test of Structural Reasoning: scores

	Version 1	Version 2
Mean	21,19	21,44
Standard Error	0,19	0,20
Median	23	24
Mode	30	30
Standard Deviation	8,13	7,95
Sample Variance	66,14	63,27
Kurtosis	-0,31	-0,37
Skewness	-0,80	-0,81
Range	30	30
Minimum	0	0
Maximum	30	30
Count	1880	1507



Data comparing pre- and post-training results

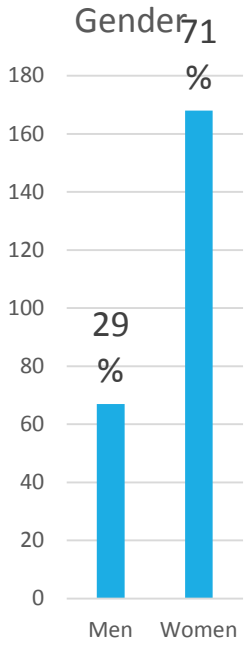
Integration training in 8 schools in capital area

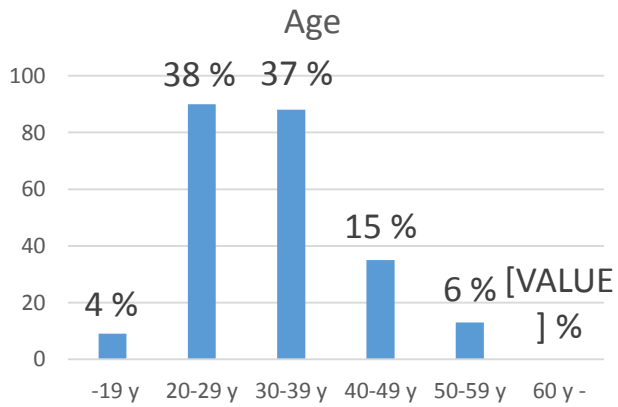
- 3 different kinds of study paths: fast, medium & slow
- Proficiency tests developed at Testipiste
- CEFR levels in speaking, writing, reading & listening comprehension

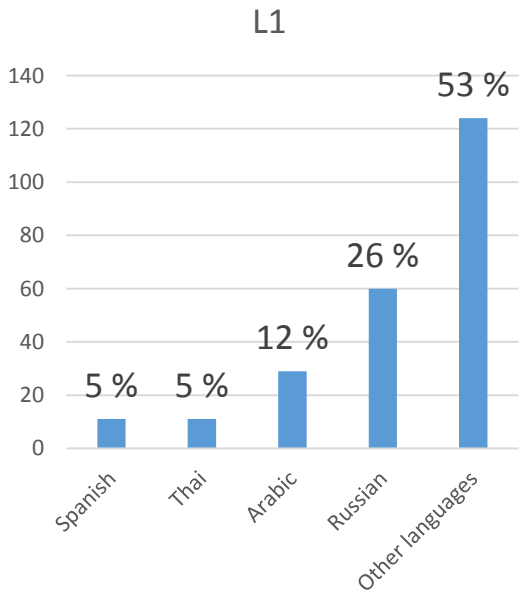
Initial testing at Testipiste

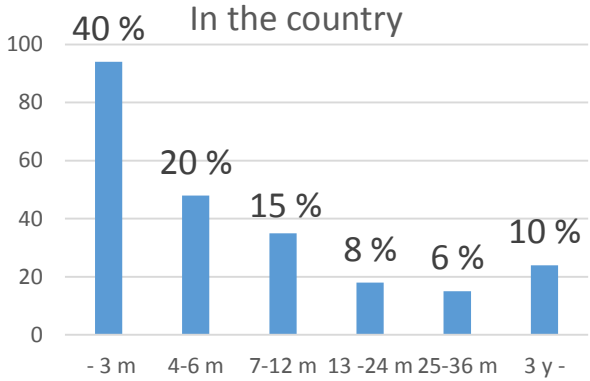
- Background information and test results
- Taken to this study if…
  - Finnish skills in initial testing 0 or in speaking pre-A1
  - No previous studies in Finnish
  - N = 235

Data comparing pre- and post-training results

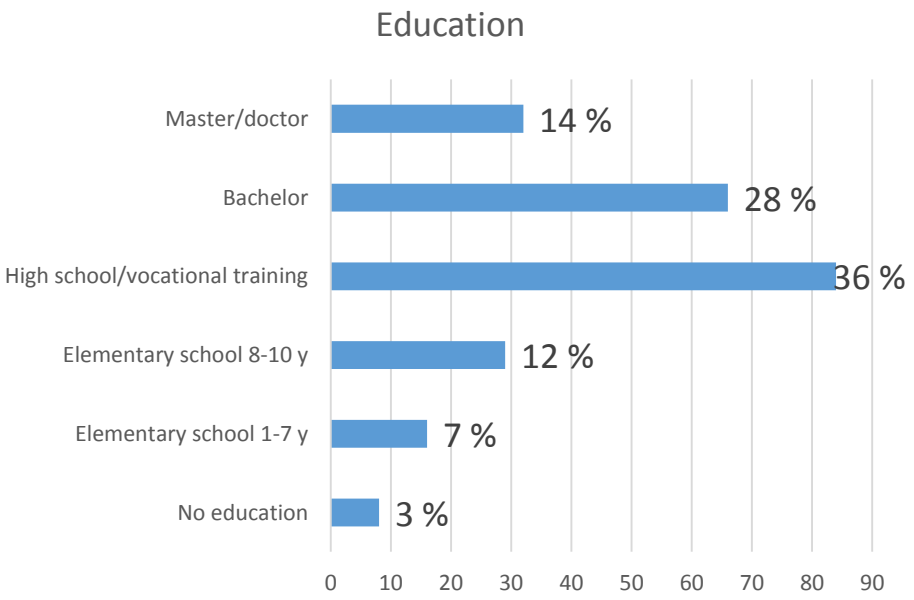


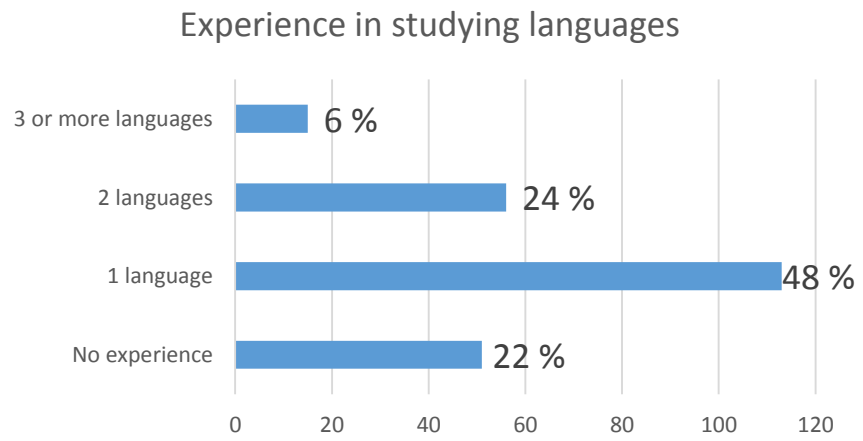


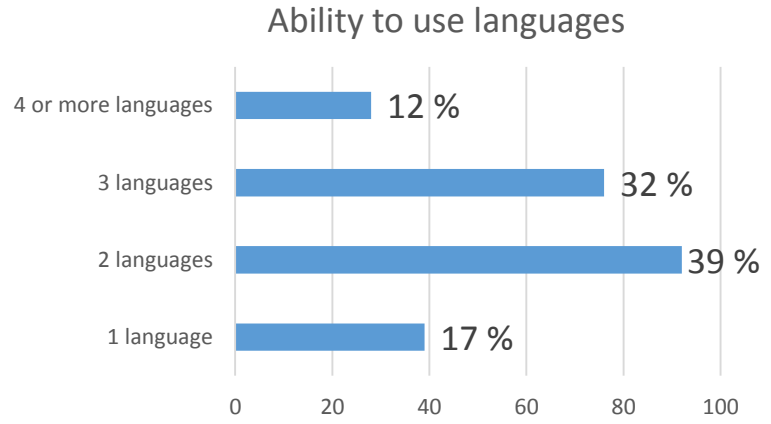




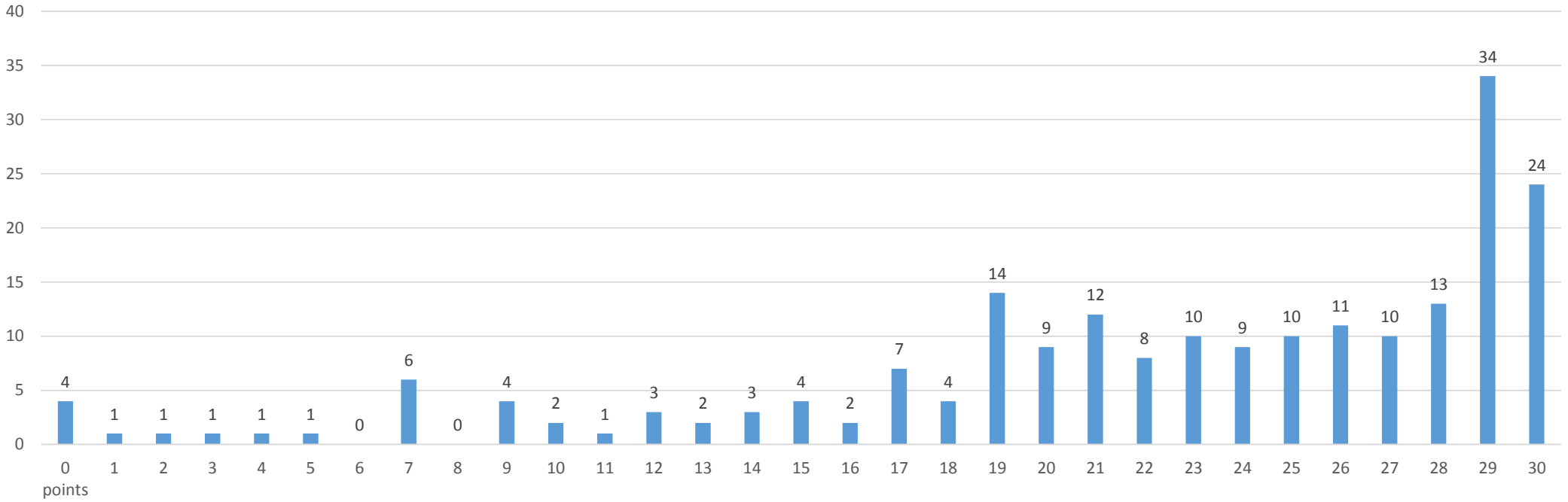
Data comparing pre- and post-training results





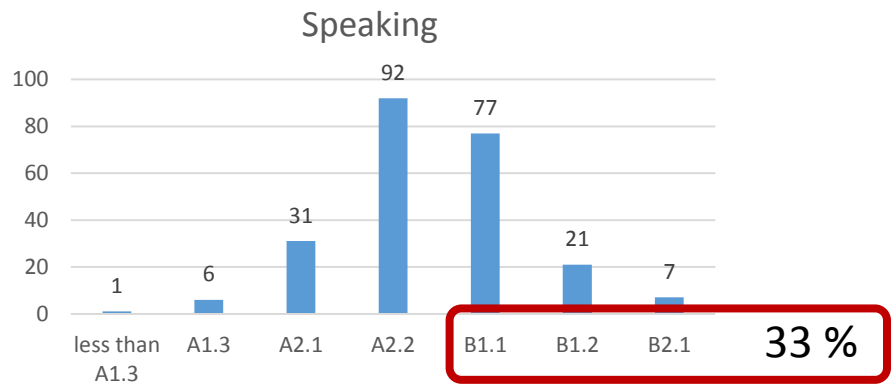


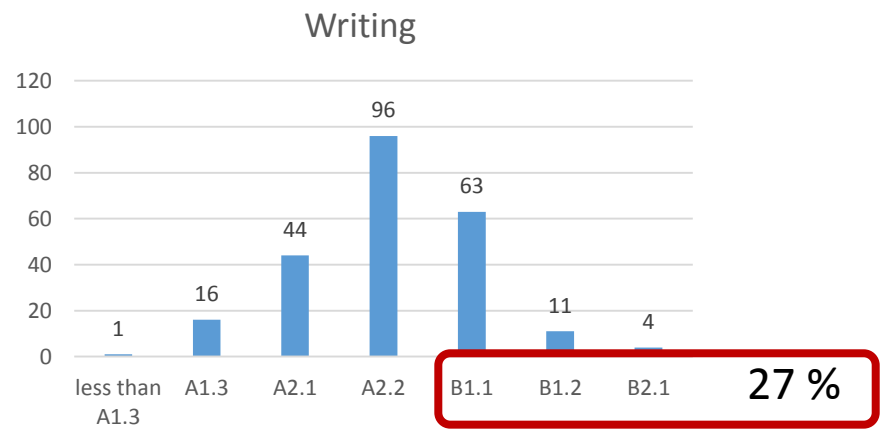
Pre-training results: structural reasoning

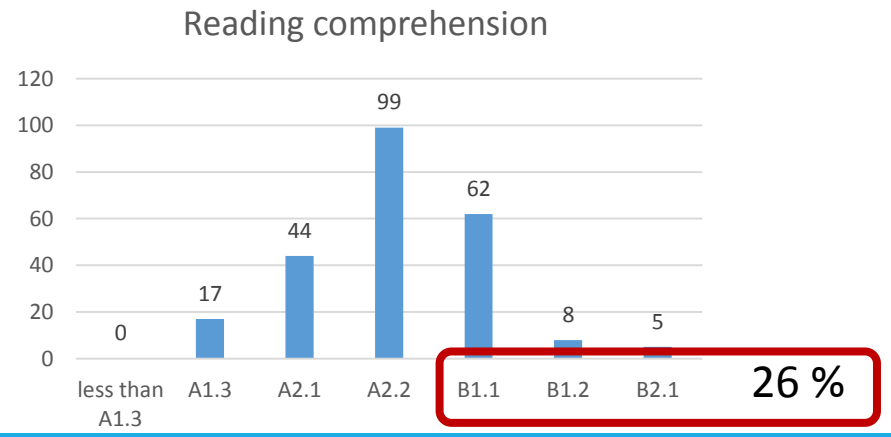


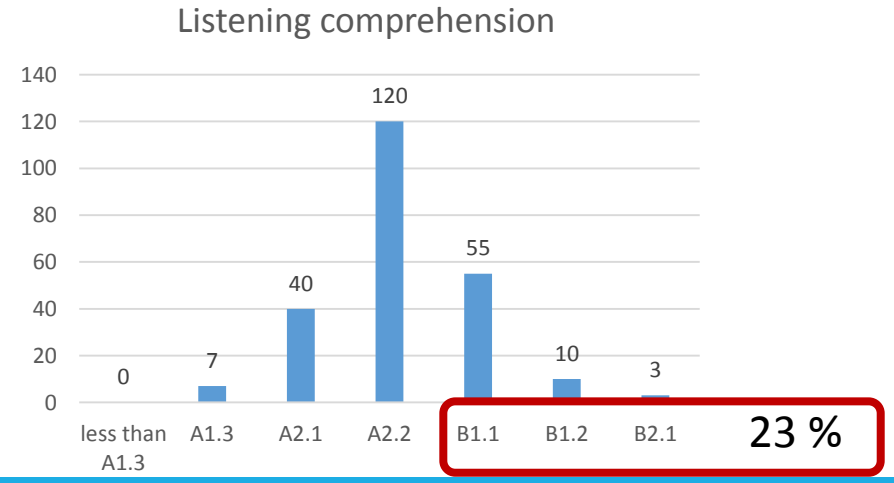
Mean	22.31	Mode	29
S. E. Mean	.51	Median	24
SD	7.47	Range	0 … 30

Post-training results: Finnish skills









Analyses and results

Correlations

		Speaking LPU	Writing LKI	Reading LLY	Listening LKY
RH	Pearson Correlation	.34	.42	.50	.32
	Sig. (2-tailed)	.000	.000	.000	.000
	N	211	211	211	211

Reading comprehension: shared variance 25 %

Analyses and results

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
LPU	Speaking	.49	5	205	.785
LKI	Writing	.50	5	205	.773
LLY	Reading	1.64	5	205	.150
LKY	Listening	2.29	5	205	.047

Critical F value 3.11 (.00)

r .374 … .574
Reading comprehension: 33 % of the total variance explained by the test of structural reasoning

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
LPU Spe	Between Groups	27.61	5	5.52	5.75	.000
	Within Groups	197.00	205	.96		
	Total	224.61	210			
LKI Wri	Between Groups	45.23	5	9.05	10.34	.000
	Within Groups	179.28	205	.87		
	Total	224.51	210			
LLY Rea	Between Groups	51.63	5	10.33	13.49	.000
	Within Groups	156.88	205	.77		
	Total	208.51	210			
LKY Lis	Between Groups	18.06	5	3.61	5.02	.000
	Within Groups	147.58	205	.72		
	Total	165.64	210			

Analyses and results

Chi-square tests. Speaking

Statistic	Value	df	Asymp. Sig. (2-tailed)
Pearson Chi-Square	49.39	30	.014
Likelihood Ratio	49.10	30	.015
Linear-by-Linear Association	23.37	1	.000
N of Valid Cases	211		

Chi-square tests. Reading

Statistic	Value	df	Asymp. Sig. (2-tailed)
Pearson Chi-Square	85.66	25	.000
Likelihood Ratio	83.24	25	.000
Linear-by-Linear Association	48.86	1	.000
N of Valid Cases	211		

Chi-square tests. Writing

Statistic	Value	df	Asymp. Sig. (2-tailed)
Pearson Chi-Square	69.62	30	.000
Likelihood Ratio	69.39	30	.000
Linear-by-Linear Association	37.52	1	.000
N of Valid Cases	211		

Chi-square tests. Listening

Statistic	Value	df	Asymp. Sig. (2-tailed)
Pearson Chi-Square	43.83	25	.011
Likelihood Ratio	48.66	25	.003
Linear-by-Linear Association	20.80	1	.000
N of Valid Cases	211		

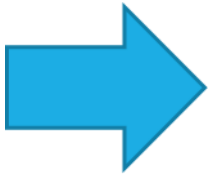
There is a significant association between language outcomes and the test results.

Conclusions?

Educational background ?

Technical reading and writing skills before training ?

Weaker results in the Test of Structural Reasoning
~ lower proficiency levels in the end of training
especially in reading comprehension and writing



More support and more training for the weaker test takers?

Study path ?

Ability to use other languages ?

Gender ?

Amount of studied languages ?

Age ?