

ALWAYS LEARNING

Paired speaking tests: are they fair?

Glyn Jones

Overview

- The test
- The sample
- The study
- Results

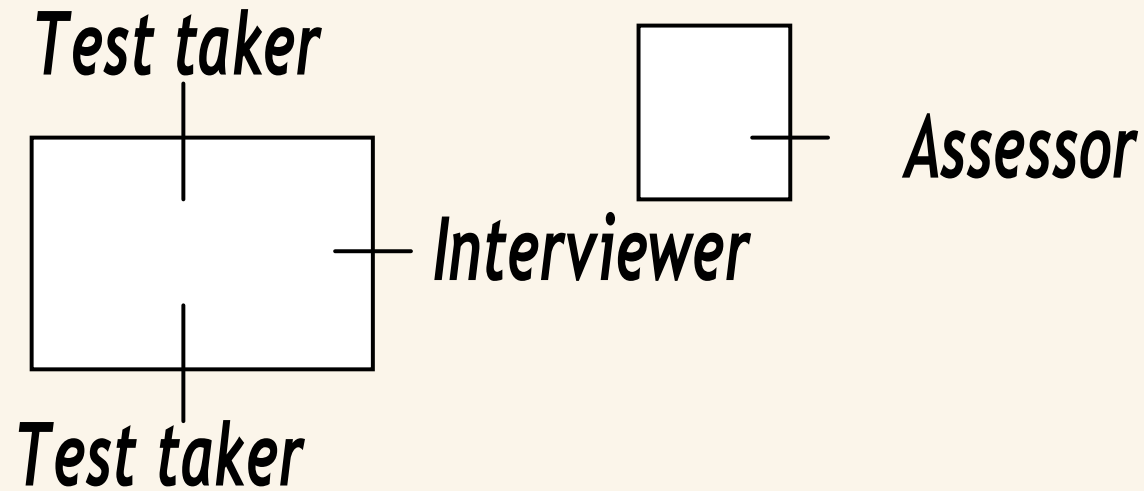
The test

London Test of English (LTE)

(Now superseded by Pearson Test of English General (PTE General))

Level	Duration
A1	10 minutes
1	10 minutes
2	10 minutes
3	12 minutes
4	15 minutes
5	15 minutes

Test format: room layout



Test format: three phases

Phase 1 Introduction

The test takers exchange personal information with the interviewer and with each other.

Phase 2: Theme/Scenario

In this phase the theme of the oral test is introduced. The test takers conduct a conversation based on a picture chosen by the interviewer from a selection included in the oral examination paper.

Phase 3 Development

The test takers conduct a conversation in greater depth based on another picture and one or more further discussion points (Levels 2 to 5) chosen by the interviewer.

Scoring

0 – 5 points for each of 5 traits (max 25 points)

- Effective communication
- Interaction
- Vocabulary
- Grammar
- Pronunciation

The sample

N = 828

Female: 562 ; Male: 266 ; 168 x M/F dyads

Level	N
A1	620
1	96
2	96
3	16
Total	828

Mean scores

Level	Mean score (out of 25)	SD
A1	16.15	5.08
1	17.36	5.14
2	17.12	4.19
3	16.40	4.15
Overall	16.40	4.92

Mean scores by gender

gender		N	Mean	Std. Deviation
Speaking Score	M	266	16.55	4.804
	F	562	16.33	4.984

Mean scores by gender

mix	Mean	N	Std. Deviation
FF	16.42	394	4.971
FM	16.13	168	5.022
MF	16.61	168	4.895
MM	16.43	98	4.666
Total	16.40	828	4.925

The study

- **Correlations:**
scores with partner scores
- **Multiple regression**
scores by partner scores, raters and interlocutors
- **Independent rating of edited files**

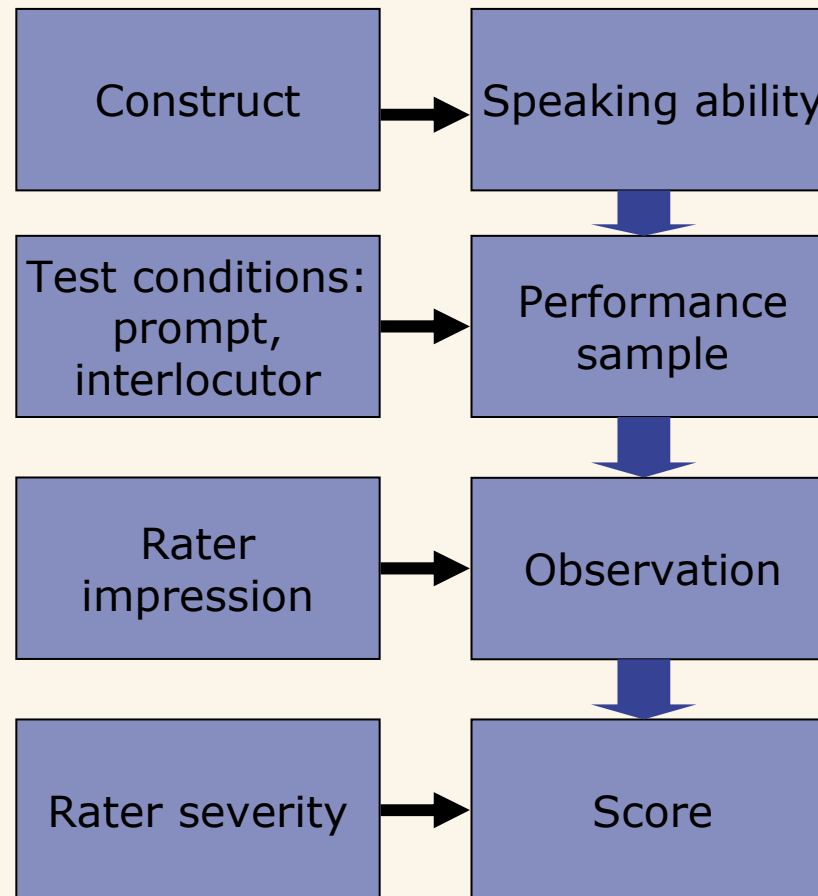
Correlation with partner's score

		Speaking score	Partner's speaking score
S_Score	Pearson Correlation	1	.471**
	Sig. (1-tailed)		.000
	N	414	414
p_s_score	Pearson Correlation	.471**	1
	Sig. (1-tailed)	.000	
	N	414	414

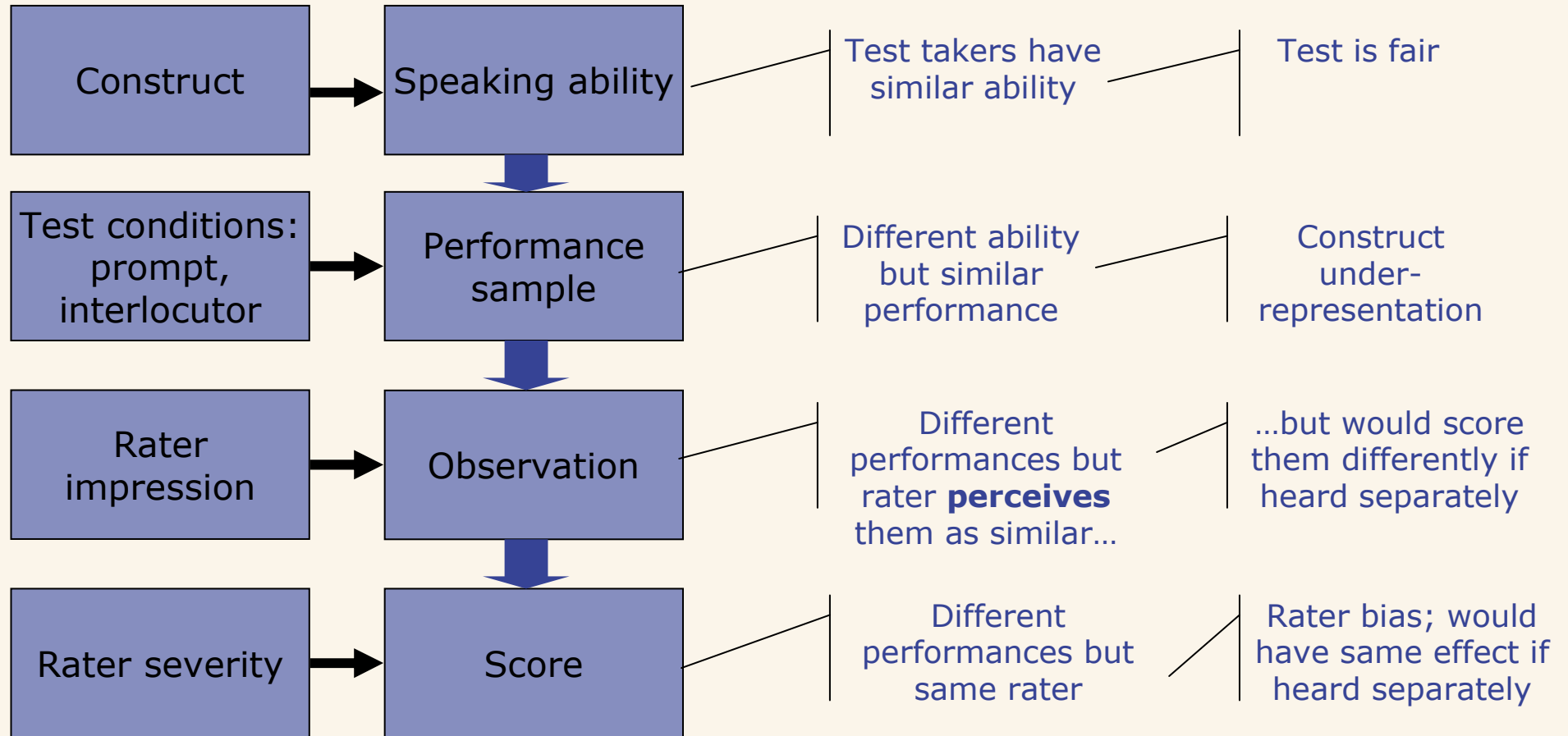
Correlations score/partner score

sample	Pearson r score/partner score	N
Level A1	.446	310
Level 1	.364	48
Level 2	.628	48
overall	.471	414

Sources of variance



Sources of variance



Why do scores tend to converge with partner scores?

Hypothesis	Convergence attributable to...	Plausible interpretation
1	Similar speaking ability; the test takers are not paired randomly	The assessment is valid
2	Similar performance; true speaking abilities differ but performance in the test is influenced by partner's performance	The score is a valid reflection of the performance as observed, but the performance is not representative of true ability; construct under-representation
3	Same occasion	When scoring one test taker, the rater is influenced by his/her impression of the other (a similar bias would NOT necessarily be expected if the same rater scored the test taker separately)
4	Same rater	Scores are subject to rater bias (but similar bias might be expected if the same rater scored the test takers separately)

Correlations score/partner score

sample	Pearson r speaking score/partner speaking score	Pearson r speaking score/partner LRW score	N
Level A1	.446	.236	310
Level 1	.364	.147	48
Level 2	.628	.214	48
overall	.471	.240	414

Multiple regression: Speaking score by partner speaking score, partner LRW score and rater ID

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.471 ^a	.222	.220	4.190
2	.488 ^b	.238	.234	4.151
3	.506 ^c	.256	.250	4.107
4	.517 ^d	.268	.261	4.079
5	.531 ^e	.282	.273	4.043

Multiple regression: Speaking score by partner speaking score, partner LRW score and rater ID

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.260	.693		13.370	.000
	p_s_score	.437	.040	.471	10.828	.000
5	(Constant)	10.608	.720		14.728	.000
	p_s_score	.357	.041	.385	8.637	.000
	A100589	2.663	.883	.129	3.017	.003
	A510895	2.616	.887	.126	2.948	.003
	A975750	-2.048	.701	-.125	-2.920	.004
	A980868	-2.531	.879	-.125	-2.881	.004

a. Dependent Variable: S_Score

Multiple regression: Speaking score by partner speaking score, partner LRW score and interlocutor ID

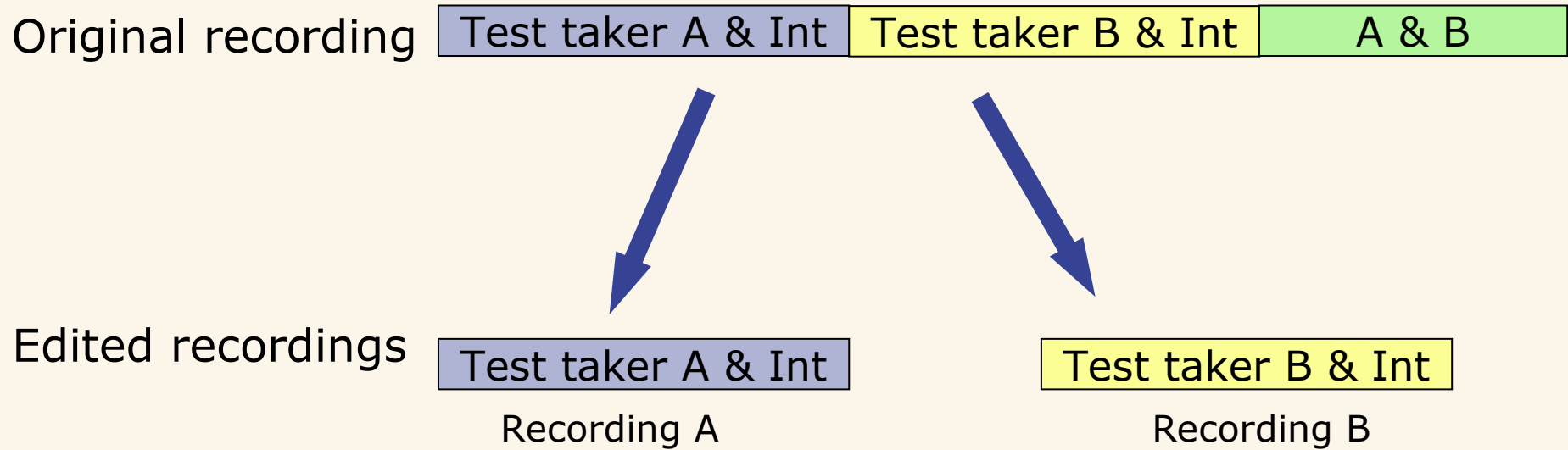
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.471 ^a	.222	.220	4.190
2	.488 ^b	.238	.234	4.151
3	.506 ^c	.256	.250	4.107
4	.514 ^d	.264	.257	4.089
5	.522 ^e	.272	.263	4.071

Multiple regression: Speaking score by partner speaking score, partner LRW score and interlocutor ID

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.260	.693		13.370	.000
	p_s_score	.437	.040	.471	10.828	.000
5	(Constant)	10.427	.728		14.331	.000
	p_s_score	.363	.042	.391	8.699	.000
	I986319	2.732	.888	.132	3.076	.002
	I986564	2.681	.893	.130	3.002	.003
	I975737	-2.173	.915	-.103	-2.375	.018
	I100587	-1.563	.721	-.094	-2.170	.031

a. Dependent Variable: S_Score

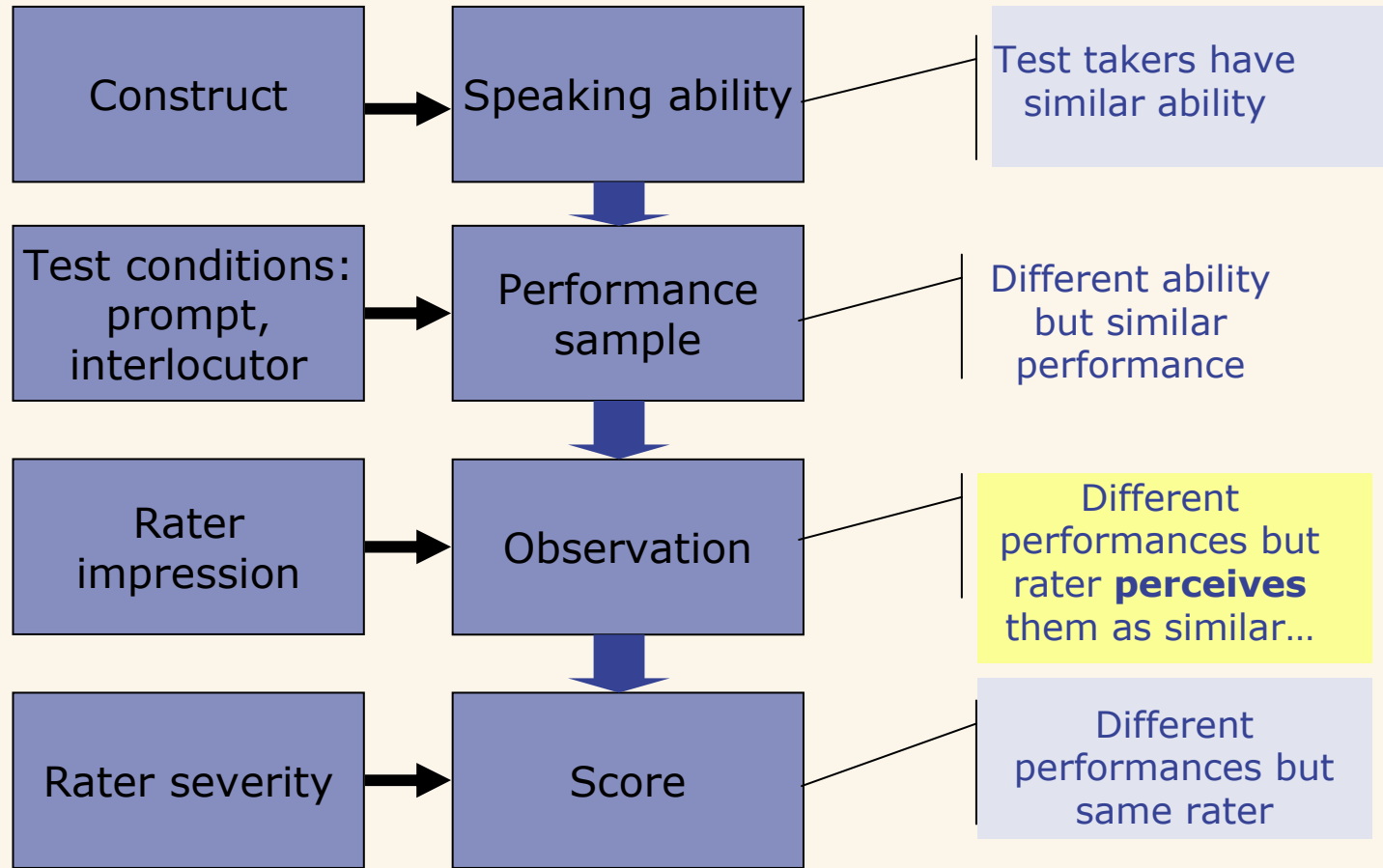
Independent rating of edited audio files



Correlation between speaking score and partner speaking score when rated separately: N = 140

		Speaking score (paired)	Partner Speaking score (paired)	Speaking score (independent rating)	Partner speaking score (independent rating)
S_Score	Pearson Correlation Sig. (1-tailed)	1	.614** .000	.619** .000	.389** .000
p_s_score	Pearson Correlation Sig. (1-tailed)	.614** .000	1	.316** .004	.633** .000
external av	Pearson Correlation Sig. (1-tailed)	.619** .000	.316** .004	1	.374** .001
p_external av	Pearson Correlation Sig. (1-tailed)	.389** .000	.633** .000	.374** .001	1

Sources of variance



glyn.jones@pearson.com