



Language Proficiency Testing and Fairness

Frans van der Slik
Radboud University
The Netherlands
f.v.d.slik@let.ru.nl



This is a Language



Tester's Paradise!



- Mastering an L2 is of vital importance in multilingual countries like South-Africa.
- But when does one master a second language sufficiently and how is that recognized?
- This is where language testing often appears on the screen.



- **Common European Framework of Reference for Languages (CEFR)**
 - can be used as a kind of metric to evaluate language proficiency tests
- **Three levels of mastery:**
 - **Basic user A**
 - A1 Breakthrough
 - A2 Waystage
 - **Independent user B**
 - B1 Threshold
 - B2 Vantage
 - **Proficient user C**
 - C1 Effective operational mastery
 - C2 Mastery



• CEFR Can-do Statements (Example)

B2 | Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

Retrieved from:

www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf



- **Fairness** is a central issue in language testing (McNamara 2000)
- In order to make fair decisions, a test has to be both **reliable**, that is: it measures consistently, and it has to be **valid**, that is: it measures what it is intended to measure



- **Reliability**
- A test is said to be reliable if the rank order of the testees would be the same when the test is repeated a large number of times (parallel tests)
- A test is also said to be reliable if it depicts the degree to which the observed scores represent the “true” scores



- **Validity**

- Content or Face validity
- Construct validity
- Concurrent validity
- Predictive validity



- **Illustration**

- **Test of Academic Literacy Levels**

- Section 1: Scrambled text
- Section 2: Knowledge of academic vocabulary
- Section 3: Interpreting graphs and visual information
- Section 4: Text types
- Section 5: Understanding texts
- Section 6: Text editing

• Study 1: Reliability concerns

Table 1: Descriptive statistics of the Afrikaans version of the academic literacy test

	UP	US	NW
N	2,701	1,702	2,521
Cronbach's alpha	.81	.91	.83
GLB	.88	.94	.89
Average <i>Rit</i>	.31	.43	.33
Cut-off point	60.5	50.5	55.5



Table 2: Potential misclassifications on the Afrikaans version of the academic literacy test. In italics the corresponding intervals (and in terms of standard deviations) around the cut-off points.

	UP	US	NW
Alpha based:			
Correlation between test and hypothetical parallel test	415 (15.4%) <i>57-63 (.30)</i>	192 (11.3%) <i>46-55 (.26)</i>	414 (16.4) <i>52-59 (.25)</i>
Correlation between observed and “true” scores	300 (11.1%) <i>58-62 (.22)</i>	137 (8.1%) <i>47-54 (.21)</i>	298 (11.8) <i>53-58 (.20)</i>
GLB based:			
Correlation between test and hypothetical parallel test	349 (12.9%) <i>58-62 (.22)</i>	157 (9.2%) <i>47-54 (.21)</i>	343 (13.6) <i>53-58 (.20)</i>
Correlation between observed and “true” scores	250 (9.3%) <i>59-61 (.15)</i>	112 (6.6%) <i>48-53 (.15)</i>	245 (9.7) <i>54-57 (.13)</i>



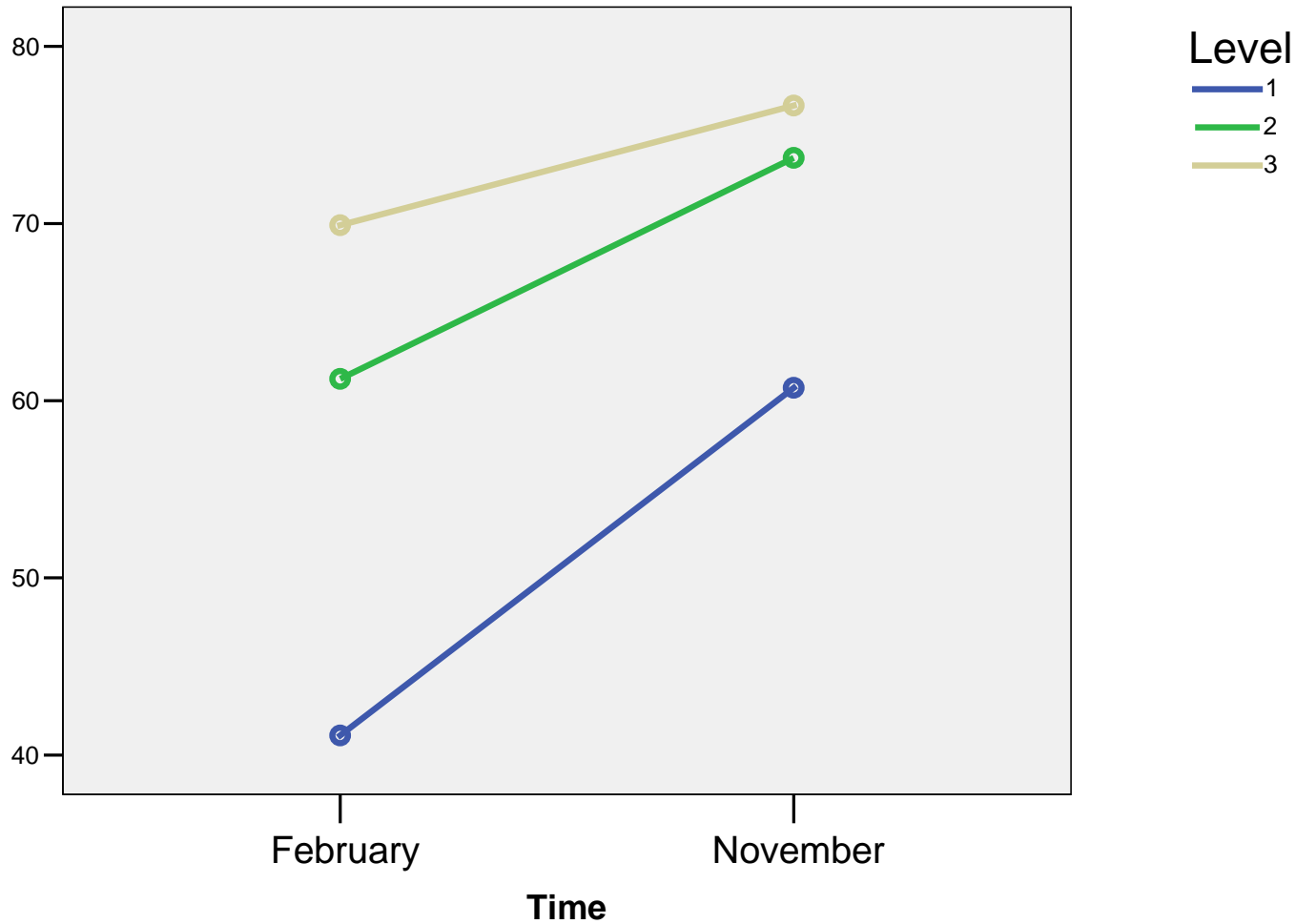
• Study 2: Validity Concerns

Table 3: Descriptive statistics of the English version of the academic literacy test in February and November 2005 by initial level of academic literacy in February 2005 and by mother tongue

		Mother tongue English or Afrikaans			Mother tongue Other African Language or Other Language			Total		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
February	Level 1	43.3	8.4	22	38.9	10.1	275	39.2	10.1	297
	Level 2	61.8	4.3	84	60.8	4.7	182	61.1	4.6	266
	Level 3	69.9	.7	17	69.9	.8	30	69.9	.8	47
	Total	59.6	9.5	123	49.0	14.2	487	51.1	14.1	610
November	Level 1	66.7	13.7	22	54.7	12.8	275	55.6	13.2	297
	Level 2	74.7	9.1	84	72.5	10.3	182	73.2	10.0	266
	Level 3	77.4	7.4	17	75.0	8.1	30	76.5	7.8	47
	Total	73.6	10.4	123	62.7	14.8	487	64.9	14.7	610



Figure 1: Improvement of academic literacy over time for different levels of academic literacy.





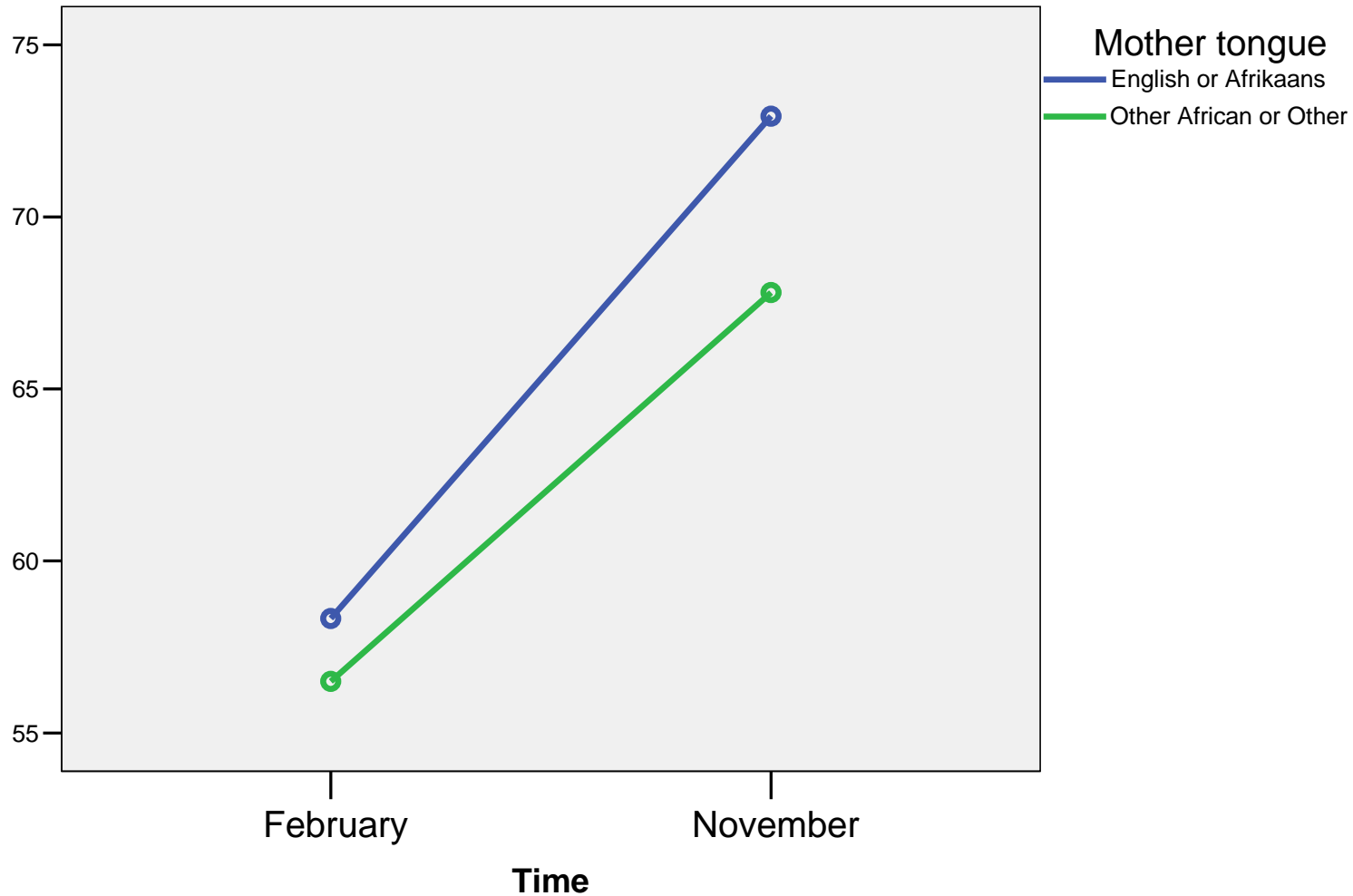
- **What caused the improvement?**

- Testing effect
- Maturing effect
- Tutoring effect
 - Matthew effect*

* Matthew: 25: 29: For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath.



Figure 2: Improvement of academic literacy over time for different mother tongues.





- **Item Response Theory**

- Keeps test difficulty constant across tests
- Ensures fairness of a decision based on the outcomes of a test
- Of vital importance in case of high-stake tests
- This is why fairness has to remain a central issue in testing

