

Improving the BECC Bunkyo English Tests Searching for Validity



What?

A three-stage process of analysis combining Rasch Analysis, Excel Analysis, and Text Analysis.

Why?

“Validity is the most fundamental consideration in developing and evaluating tests” (APA, AERA, NCME, 2014)

“The key criterion driving assessment at Cambridge Assessment is validity” (Shaw, 2020)

1. Rasch Analysis

- Person / Item Variables
- Item Correlation
- Item Statistics

2. Excel Analysis

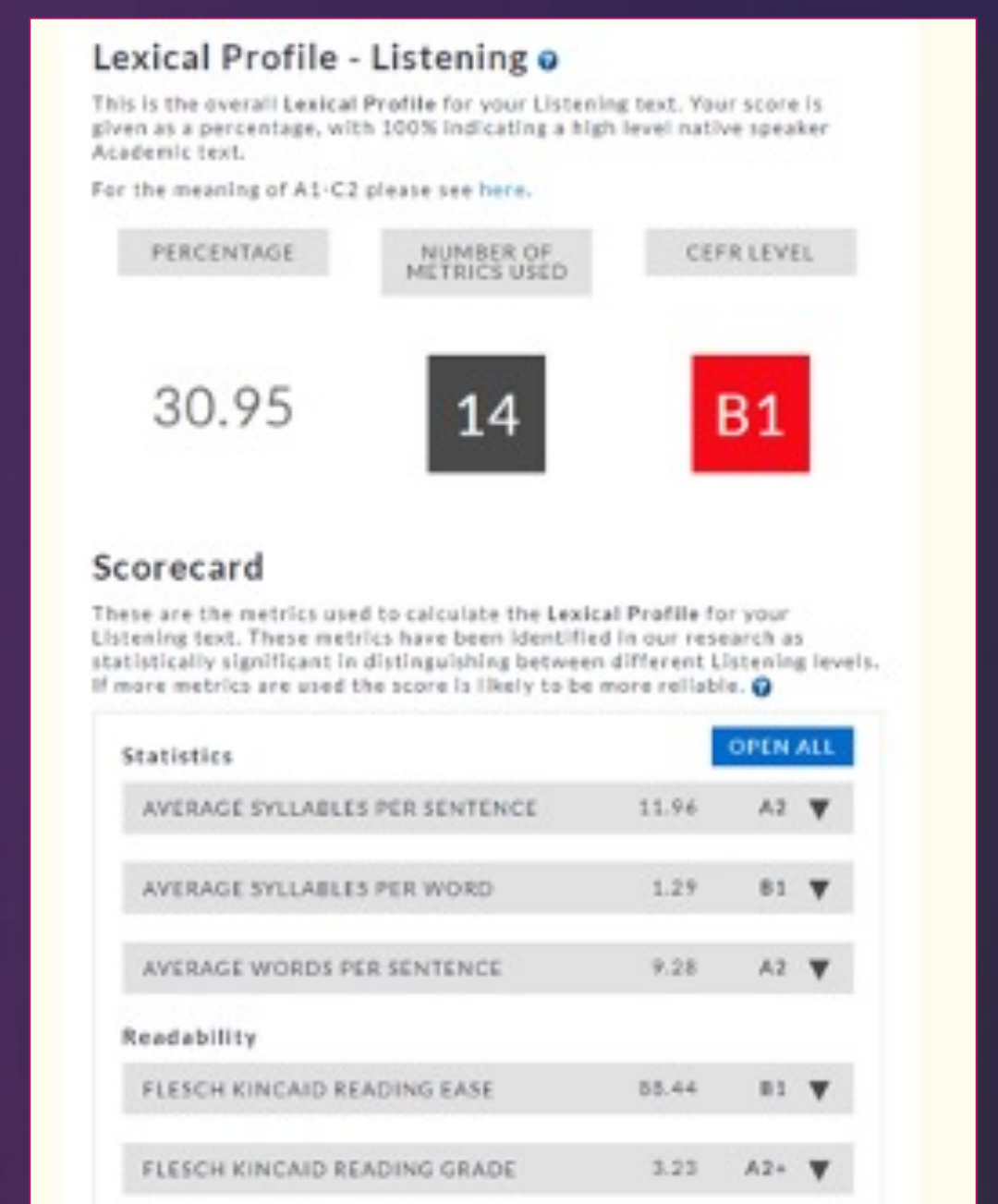
- The Excel BET Database
- Excel Question Analysis
- Excel Microanalysis
- Excel Testlet Analysis

Section	Section	Overall	Difficulty Ranking	Low Stream	High Stream	High - Low	Ability Determiner Ranking
Reading (BET)	1	65.6%	14	53.4%	73.1%	19.7%	7
	2	65.5%	13	51.9%	74.3%	22.4%	3
	3	56.6%	7	43.6%	64.2%	20.6%	5
	4	62.9%	10	46.6%	72.6%	26.0%	2
	5	64.8%	11	51.9%	72.9%	21.0%	4
	6	54.7%	4	43.7%	61.7%	17.9%	8
Listening (BET)	7	43.7%	1	26.6%	54.3%	27.7%	1
	8	58.2%	8	48.4%	64.9%	16.4%	11
	1	56.1%	6	45.7%	61.5%	15.8%	12
	2	53.6%	2	43.1%	61.0%	17.9%	9
	3	65.2%	12	52.4%	73.0%	20.5%	6
	4	55.0%	5	46.0%	60.4%	14.4%	13
5	54.7%	3	43.9%	61.0%	17.1%	10	
6	58.3%	9	51.9%	62.9%	11.0%	14	

3. Text Inspector

- BET Text Analysis
- Reading Section Analysis
- Listening Section Analysis

The GEAC committee wants to create assessments that become progressively challenging through each section of the test, therefore the CEFR level progresses from A2 to B1+.



Part	CEFR Aim	A1	A2	B1	B2	Overall CEFR	FRE/FKL
1	A2	73%	27%	-	-	NA	NA
2	A2/ A2+	61%	18%	15%	-	NA	NA
3	A2/ A2+	69%	7%	5%	-	NA	90 / 1.8
4	A2/ A2+	80%	7%	3%	-	A1+	89 / 2.3
5	A2/ B1	66%	22%	5.5%	1%	A2	72 / 6.5
6	A2/ A2+	57%	25%	11%	-	A2+	80 / 5.5
7	B1/ B1+	63%	17.5%	9%	-	A2	82 / 4.4
8	B1/ B1+	60%	14%	7%	-	B1	65 / 8.5

Part	CEFR Aim	A1	A2	B1	B2	Overall CEFR	FRE/FKL
1	B1	58%	17%	8%	1.5%	B1	83 / 4.0
2	A2	66%	14%	3.5%	1%	B1+	85 / 3.25
3	A2	73%	10%	1%	-	A2+	86 / 3.0
4	A2	62%	14%	6%	5%	B2	83 / 3.85
5	B1/ B1+	57%	21%	12%	2%	B2	83 / 3.85
6	B1/ B1+	70%	15%	7%	1%	B1	83 / 4.26

Conclusions

“No test is valid for all purposes or in all situations or for all groups of individuals.” (Shaw 2020)

All we can do is try to provide evidence of having performed due process to satisfy the following stages of construct validity:

- The contents of the tests match the contents of the curriculum
- Students answer the questions in the manner intended, and their answers to different questions relate in an expected way
- Users can interpret results in the manner intended

