

Introduction

BME Language Examination Centre (Hungary)

- examinations accredited by the state
- five languages: English, German, French, Italian and Spanish
- three levels: B1, B2, C1
- monolingual, four skills
- over 300 examiners in Hungary

Annual standardisation further training for oral examiners

Until 2016: traditional, face-to-face mode
Since 2017: online mode

Two sections of the online mode:

- Examiners complete a test on their knowledge of the CEFR levels
- Examiners evaluate examinees' oral language performance recorded on audio files

Research questions

- (RQ1) Are there any differences between the way examiners of different languages interpret the CEFR scales?
- (RQ2) Are there any differences between the way examiners of different languages interpret the criteria of the analytical rating scale used for assessing examinees' oral language performance?
- (RQ3) What are the difficulties of turning from traditional to online oral exam standardisation sessions?

Participants

Examiners	2017	2018
English	174	176
German	80	91
French	35	33
Italian	19	15
Spanish	11	12
Total	319	327

Research instruments

- A 35-item CEFR test – 12 B1, 12 B2 and 11 C1 descriptors to be categorized by examiners
- Examinees' oral language performances on the Speaking test recorded on audio files evaluated by examiners with an analytical rating scale
Speaking test: 1. Topic discussion
2. Monologue 3. Situation
Analytical rating scale: five criteria: Communicative value; Expressiveness; Grammatical accuracy; Discourse management; Pronunciation, intonation
- An online questionnaire to collect examiners' feedback on the online training

(RQ1) Examiners' performance on the CEFR test

(1) Descriptive statistics 2017:

2017	Eng.	Ger.	Fren.	Ital.	Spa.
Mean	28,65	27,88	28,49	27,32	25,96
St.dev.	4,94	6,18	5,54	5,62	4,78
Mode	34	35	33	33	23
Median	29,5	29	29	28	25
Minimum	12	9	16	17	19
Maximum	35	35	35	35	35

(2) Statistical comparisons (Wilcoxon Signed Ranks Tests) of examiners' performance on the CEFR test across languages 2017:

2017	Eng - Ger	Eng - Fr	Eng - Ital	Eng - Sp	Ger - Fr
Z	-,567	-,814	-,068	-,1469	-,176
Asymp. Sig. (2-tailed)	,571	,415	,945	,142	,861

	Ger - Ital	Ger - Sp	Fr - Ital	Fr - S	Ital - Sp
Z	-,409	-,1465	-,680	-,1915	-,376
Asymp. Sig. (2-tailed)	,683	,143	,497	,056	,707

(3) Descriptive statistics 2018:

2018	Eng.	Ger.	Fren.	Ital.	Spa.
Mean	29,47	28,48	30,73	28,87	27,39
St.dev.	4,85	5,49	4,27	5,36	5,39
Mode	35	35	32	26	32
Median	31	30	32	29,5	27
Minimum	12	15	18	13	17
Maximum	35	35	35	35	35

(4) Statistical comparisons (Wilcoxon Signed Ranks Tests) of examiners' performance on the CEFR test across languages 2018:

2018	Eng - Ger	Eng - Fr	Eng - Ital	Eng - Sp	Ger - Fr
Z	-2,130	-1,428	-1,149	-,017	-2,122
Asymp. Sig. (2-tailed)	,033	,153	,251	,986	,034

	Ger - Ital	Ger - Sp	Fr - Ital	Fr - Sp	Ital - Sp
Z	-1,464	-1,592	-,926	-1,625	-,797
Asymp. Sig. (2-tailed)	,143	,111	,354	,104	,425

There was no statistically significant difference between examiners' CEFR performance across languages in 2017 and in 2018.

The only statistically significant difference at p<0.05 was found in Eng-Ger and Ger-Fr in 2018.

(5) Descriptors examiners are most familiar with:

2017: 8 descriptors matched with right level over 91% precision – 8 related to **communicative language competences**

2018: 13 descriptors matched with right level over 91% precision – 10 related to **communicative language competences**

(RQ2) Examiners' use of the analytical rating scale

(1) Standard points for examinees' oral language performance:

	English	German	French	Italian	Spanish
Maximum			60		
Passmark			36		
2017 standard	37	40	44	33	38
2018 standard	50	38	34	38	27

(2) Most severely rated criteria (based on mean):

2017	English	German	French	Italian	Spanish
1.	Grammar3	Express1	Grammar3	Express2	Grammar3
2.	Grammar2	Grammar1	Grammar2	Grammar2	Disc.man.2
3.	Disc.man.2	Com.val.1	Disc.man.1-3	Disc.man.2	Express3

2018	English	German	French	Italian	Spanish
1.	Disc.man.2	Grammar3	Grammar2	Com.val.2	Grammar1
2.	Grammar2	Express3	Grammar1	Express2	Grammar3
3.	Express2	Grammar2	Express2	Grammar2	Grammar2

(3) Most leniently rated criteria (based on mean):

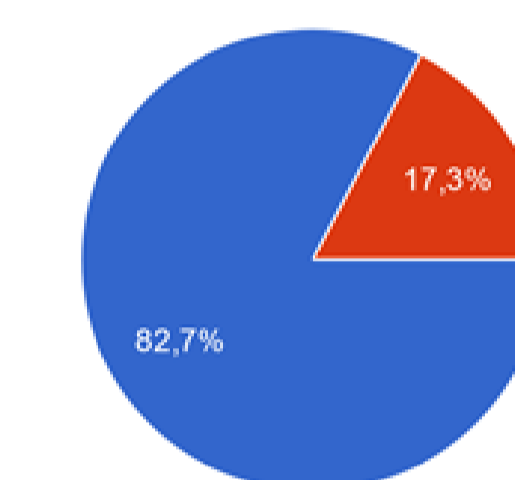
2017	English	German	French	Italian	Spanish
10.	Pronun.	Pronun.	Express2	Com.val.3	Com.val.2
11.	Com.val.3	Disc.man.2	Com.val.1	Com.val.1	Express1
12.	Com.val.1	Com.val.2	Com.val.2	Pronun.	Com.val.1

2018	English	German	French	Italian	Spanish
10.	Disc.man.1-3	Com.val.2	Disc.man.1-3	Com.val.1	Com.val.3
11.	Com.val.3	Com.val.1	Express3	Disc.man.1-3	Com.val.1
12.	Com.val.1	Pronun.	Com.val.3	Pronun.	Pronun.

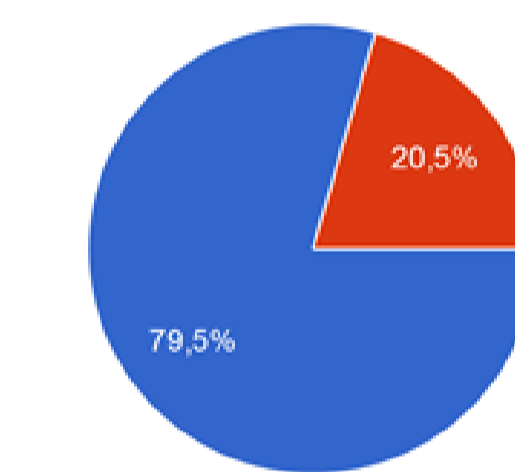
Regardless of the standard points and regardless of the languages the most severely rated criteria are related to **communicative language competences**.

(RQ3) Examiners' feedback on online vs traditional training

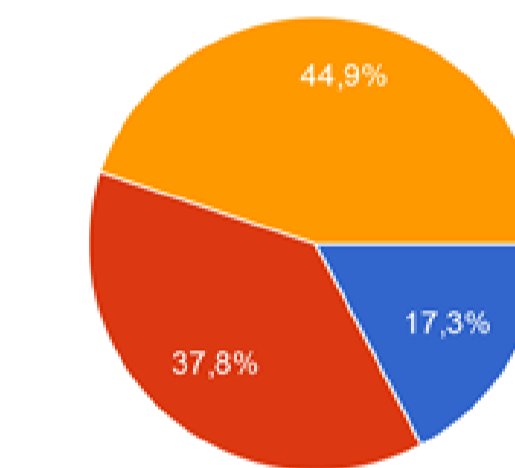
N: 156



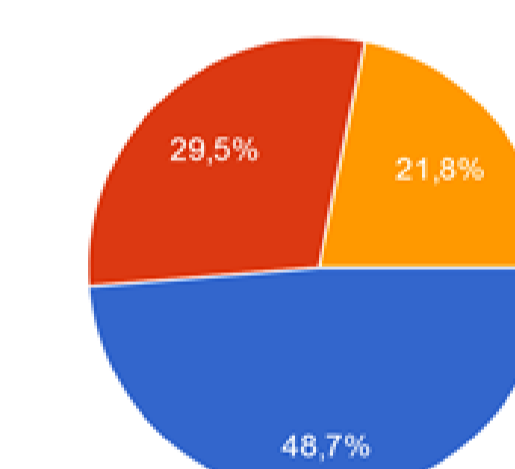
During the traditional standardisation retrainings the completion of a CEFR test was not a requirement, while during the online mode it is. Do you find it useful that a CEFR test is an obligatory element to successfully complete the training? **Yes No**



Who did you do the evaluation task with?
Alone With (a) colleague(s)



Which training mode do you prefer?
Traditional Online Neither is better



Do you agree with the assumption that the online mode is not really a training but rather a test?
Yes No Perhaps

Reasons for the online mode:

Professional reasons	8,4%
Practical reasons (comfortable, quick, cheap etc.)	91,6%

Reasons for the traditional mode:

Professional reasons (discussion with examiners, trainers; quick feedback etc.)	97,1%
Practical reasons	2,9%

Conclusions and implications

- There is no difference between examiners of different languages in terms of their familiarity with the CEFR descriptors.
- Examiners are most familiar with the CEFR descriptors related to communicative language competences.
- The most severely rated criteria are the ones related to communicative language competences regardless of the standard points awarded to oral performances and regardless of the languages.
- The online mode is more practical, flexible and comfortable but lacks immediate professional help.
- Advantages of online and traditional modes need to be combined: more detailed feedback has to be provided for examiners to account for the standard points.