How prejudices affect rating (case study Slovene)

8 November 2019

Petra Likar Stanovnik
Ethnic prejudice is an antipathy based upon a faulty and inflexible generalization. It may be felt or expressed. It may be directed toward a group as a whole, or toward an individual because he is a member of that group.

(Allport 1966, p. 9)
Fairness
Procedural fairness can be viewed as a lack of bias for or against any individual or group . . .

(Kane 2010, p. 178)
bias → reliability → test quality
sources → bias

prejudice
How prejudices affect rating (case study Slovene)

8 November 2019

Petra Likar Stanovnik
Slovene
cannot be learned

Slovene

less valuable
Slovene language exams

- Basic level (A2/B1); 92%
- Intermediate level (B2); 4%
- Advanced level (C1/C2); 4%

2018 data

Basic Level Exam
Candidates
Results
Basic Level Exam
A2/B1 CEFR

60% 85%

Reading
Listening
Speaking
Writing
Candidates

Speakers of other languages; 14%

Speakers of closely related languages; 86%

2018 data
Test results

2018 data
2018 data
How prejudices affect rating (case study Slovene)

8 November 2019

Petra Likar Stanovnik
Research

Rating analysis

Attitudes on rating criteria
Test

standardized statistical analysis

classical test analysis

IRT analysis
Test takers

low motivation

↑

lack of time
similarity of languages
low education
...

(Ferbežar 2012)
Raters

2) Attitudes on rating criteria

1) Rating analysis

Research: 2018; Ferbežar, Likar Stanovnik
Test takers

Two guided tasks:
- a practical text in a hypothetical situation
- a reply to a short letter

• 40-50 words each
Raters

Rating:
- analytical scale
- categories: text content, vocabulary, accuracy, coherence and style
- 10 points per text
- 20 points for writing subtest:
  12+ points = A2
  17+ points = B1
External raters (34)

1st rating

Centre’s raters (7)

2nd rating

3rd rating 18%

Results

final score 82%
Difference:
- number of points awarded (3+)
- pass/fail (below A2/A2)
- level (A2/B1)
Difference in scoring
81 ratings (12%)

- Speakers of other languages:
  - 1st rating more lenient: 12
  - 1st rating more severe: 2

- Speakers of South Slavic languages:
  - 1st rating more lenient: 39
  - 1st rating more severe: 28
Speakers of South Slavic languages

1st rating more lenient: 39
1st rating more severe: 28

Differential rater severity

7 test takers (1%): 1st rating: fail, 2nd & 3rd rating: pass
Questionnaire
April-June 2018

87 raters
External raters (68)  Centre's raters (19)
Should speakers of Slovene as a L1 be the reference for rating criteria?

68 external raters

- Native speakers know the language best and are most suited for setting the norm
- Non-native speakers will never be capable of speaking Slovene at the level of a native speaker
- Importance of social and symbolic role of the Slovene language

Yes; 14

No; 54
Should non-Slavic speakers be rated according to different criteria?

68 external raters

- Slavic speakers have the advantage of not having to put much effort into learning Slovene and can be better understood on account of linguistic proximity
- South Slavic language speakers are being privileged
- Favorable inclination towards non-Slavic speakers, because it is harder for them to learn Slovene

Yes; 15
No, but ...; 6
No; 47
Should the test taker’s “effort” or “path traveled” be taken into account when rating?

68 external raters

- Far greater efforts are required of non-Slavic speakers
- The test takers who speak closely related languages do not prepare for the exam and thus show their dismissive attitude towards the Slovenian language and culture
- South Slavic language speakers have no desire to learn Slovenian even after living in Slovenia for 10 or more years
Fairness

Measures:
- 16-hour introductory examiner and rater training
- regular training for raters
- a book of regulations on administering and rating exams
- standardization prior to rating speaking and writing performance
- rater monitoring
- regular checking of intra- and inter-rater agreement
Measures:
- 16-hour introductory examiner and rater training
- regular training for raters
- a book of regulations on administering and rating exams
- standardization prior to rating speaking and writing performance
- rater monitoring
- regular checking of intra- and inter-rater agreement

Research:
- rater bias resistant to change  
- unreliable scoring of written performance  
  (Brown 2015, Wind, Peterson 2018)
Fairness

Measures:
- 16-hour introductory examiner and rater training
- regular training for raters
- a book of regulations on administering and rating exams
- standardization prior to rating speaking and writing performance
- rater monitoring
- regular checking of intra- and inter-rater agreement

- raising raters' awareness of prejudices
- development of intercultural competence