

Developing a level-specific checklist for assessing writing

Zoltán Lukácsi

- local context
- testing tradition
- operational rating scales
- a viable alternative: the checklist
- research: development of a checklist
- research: achievements

- state-accredited language examination systems
- nationalization
- (alternatively: A-level GCSE in English)
- L2 certificates and social mobility
 - at present: university graduation
 - from 2020: condition for admission
- additional benefits

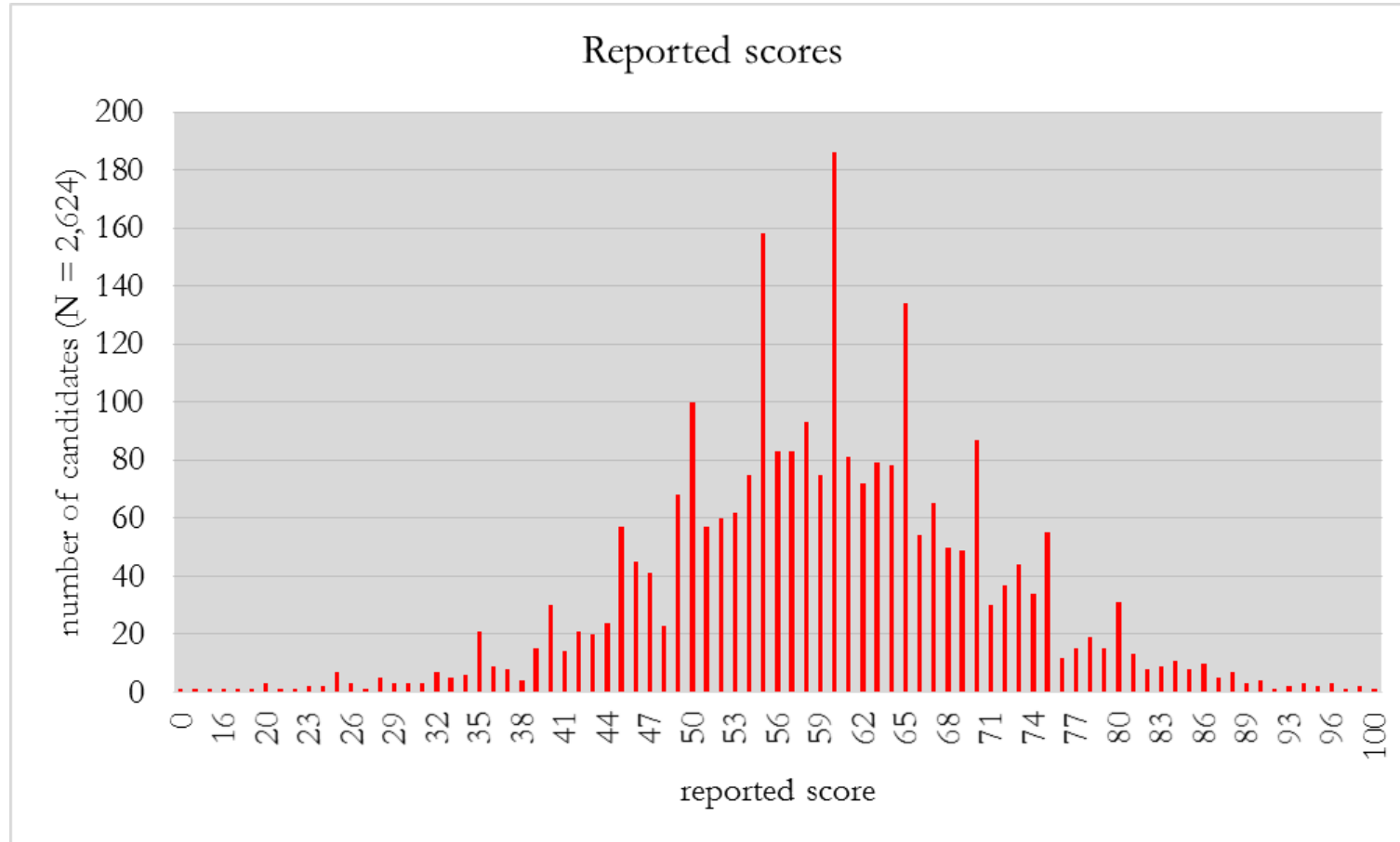
- since the early 1950s: state language exam
- paradigm shift in 1998
- Government decree 137 / 2008: **level testing**
 - oral exam (speaking and listening)
 - written exam (reading and writing) → **rating scales**

*Rating on a scale: **judging that a person is at a particular level or band on a scale made up of a number of such levels or bands** (CEFR, 2001, p. 189)*

A rating scale

Score	Task achievement	Grammatical range and accuracy
5	Task achieved at a high level Intention: Entirely clear Instructions: Completely followed Effect: A positive effect on the target reader Outcome: Sure to achieve a successful outcome Content: All relevant details included Some original ideas or presentation	Grammatical Structures: Complex Spelling : Very good Word order : Correct Punctuation: Used properly throughout Errors: Very few, none of them impedes meaning
4		
3	Task achieved, some gaps Intention: Clear in most areas Instructions: All important ones followed Effect: A generally positive effect on the reader Outcome: Likely to achieve a successful outcome Content: Many relevant details included	Grammatical Structures: Simple but mostly correct, some attempts at complex structures with some mistakes that do not impede comprehension Spelling : Some mistakes that do not impede comprehension Word order : Mostly correct Punctuation: Mostly effective Errors: Some, but do not significantly impede meaning
2		
1	Task unachieved Intention: Very unclear Instructions: Many not followed Effect: negative Outcome: Will not achieve a successful outcome Content: Omission, irrelevance	Grammatical Structures: Very simple with frequent and serious mistakes Spelling : Very poor Word order : Often wrong Punctuation: Often wrong
0	Task unattempted / partially attempted Not enough language to make an assessment, or under 20 words	Little or no evidence of grammatical knowledge of simple structures, or under 20 words

Reporting results from rating scale use



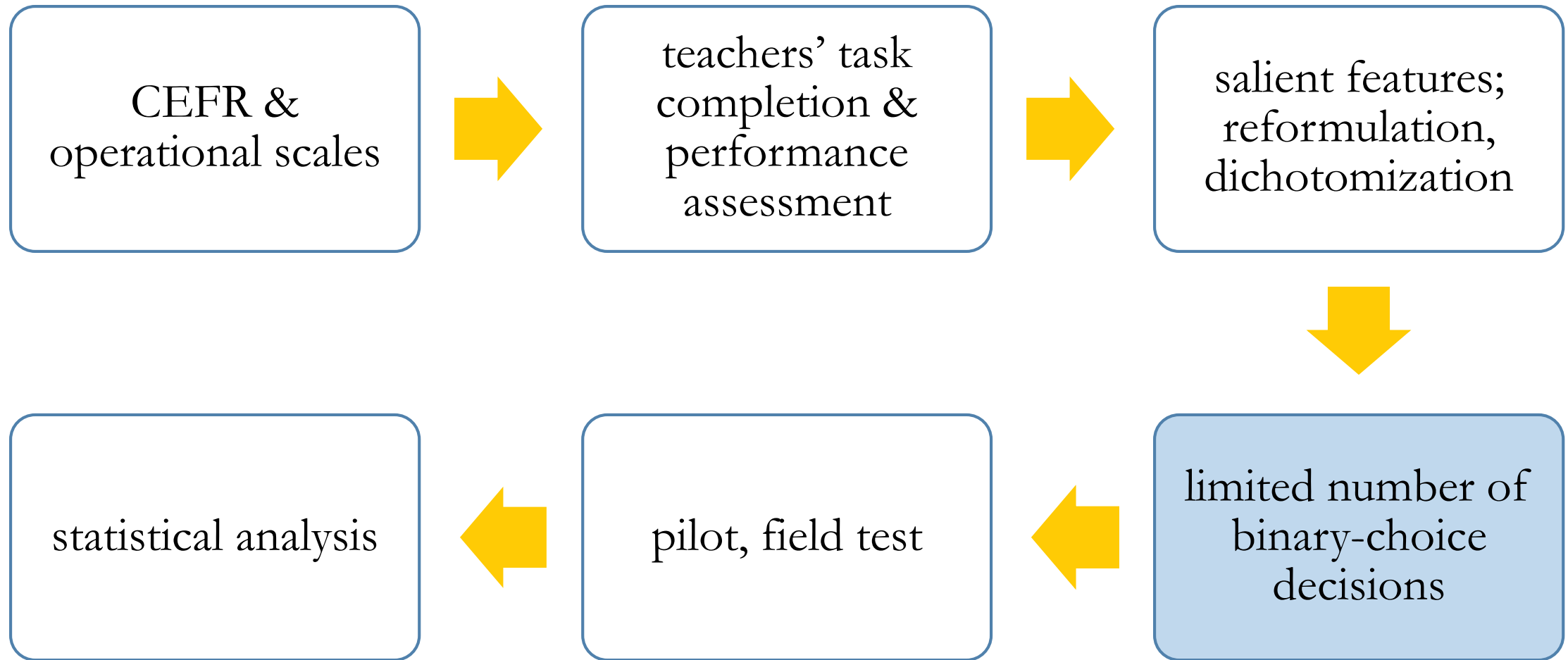
Problems with raters

If we accept the argument that opened this section that **judgements must be made by human specialists**, it should follow that **these people are able to do** what it is claimed they do. The evidence suggests that, left to their own judgements, **they cannot**.

Hamp-Lyons, L. (1990, p. 80).

- Kim, Y-H. (2011). Diagnosing EAP writing ability using the Reduced Reparameterized Unified Model. *Language Testing*, 28(4), 509-541.
- Struthers, L., Lapadat, J. C., & MacMillan, P. D. (2013). Assessing cohesion in children's writing: Development of a checklist. *Assessing Writing*, 18, 187-201.

Research plan and stages

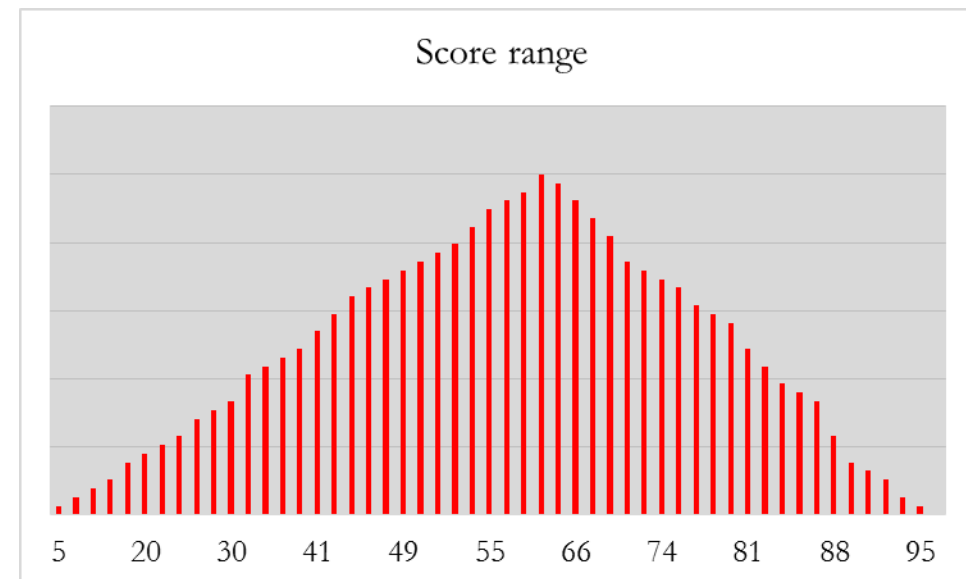


The evolution of a checklist statement

- "clear organization", "the ideas are organized into paragraphs"
- "At this point the candidate should start a new paragraph, which doesn't happen."
- Are the ideas organised into paragraphs where one paragraph stands for one idea?
- Each paragraph presents one distinct and unified idea.
- Each paragraph presents one distinct and unified idea, *to wit*:
 - Does the script follow standard paragraphing conventions?
 - Does each paragraph focus on one idea?
 - Is each body paragraph longer than a single sentence? If not, is it complex enough to realize a paragraph?

Statement	Concept check questions
1. This text is legible, i.e. the reader doesn't have to guess what the writer is trying to say.	Can you read the text without having to re-read words? Is the text legible?
2. This text follows the standard layout of an email.	Does the script have all the header fields (including "To:", "From:", and "Subject:")?
3. This text is the required length as defined by the task.	Does the body of the email have a greeting, an ending, and a signature? Is the script within the acceptable bounds as defined by the task (in this case 100-120 words)?
4. This text displays situational authenticity, i.e. the writer gives the necessary details for contextualisation, such as role, location, situation, domain, etc.	Does the writer explicitly state their role, location, situation, domain, etc.? Does the reader get detailed realistic information about the writer's context, situational underpinnings, setting, etc.?
17. This text demonstrates that the writer understands grammatical aspect in English.	Are there at least two grammatical aspects (simple, continuous, perfect) used correctly?
22. The writer can use the English lexicon to express the intended meaning instead of periphrases or non-existing terms.	Are all the words existing English items? Does the writer find the correct word instead of relying on periphrasis? Does the writer find the correct word instead of creating one?
23. Some lexical items beyond the target level are used appropriately.	Are there any lexical items beyond level B2? (use a good dictionary to check if necessary)
33. This text demonstrates that the writer can use complex linking words (such as "however", "first of all", or "what's more") meaningfully.	Are there complex linking words present? Are these used in a meaningful way? Are contrastive ties used meaningfully? Are additive ties used meaningfully?
35. Grammatical or linguistic errors in this text do not impede comprehension.	Are sequencing and enumerative ties used meaningfully? Is the script free from grammatical errors? Is the script free from other linguistic errors? OR: If errors are present, is the message still clear and obvious?

- exclude: **3. This text is the required length as defined by the task.**
- all statements proved **productive for measurement** ($0.5 \leq MS \leq 1.5$)
- only 2 statements fell outside the **reasonable fit range** ($0.7 \leq MS \leq 1.3$)
- pass vs fail classification
 - exact agreement: 89.74%
 - $\chi^2(1, N = 78) = 49.61, p < .001,$
Cramer's $V = 0.798$
- score range
 - weakest: 5, best: 96



Important advances

- broader score range, esp. at the tails
- increased transparency
- increased accountability
- more fitting to **level testing**:

Rating on a checklist: judging a person in relation to a list of points deemed to be relevant for a particular level or module (CEFR, 2001, p. 189)

Thank you

- lukacsi74@gmail.com