

Adaptive comparative judgment is a viable option for proficiency rating in second language research

What this research was about and why it is important

Proficiency is an essential construct in second language research, yet many learner production data still come with variables such as institutional status or year of study as proxies for proficiency despite the fact that these external criteria are largely regarded as unreliable. The challenge today is that researchers often do not have the financial and/or time resources to have all learner texts that they collected rated by a group of trained raters. This article proposes a new method to address this challenge by combining two techniques:

- Comparative judgment that rests on the assumption that people are able to compare two performances more easily and reliably than to assign a score to an individual performance;
- Crowdsourcing: a practice of soliciting contributions from a large group of people.

The authors illustrate this approach with an experiment that they conducted based on a set of 50 learner texts and report encouraging results.

What the researchers did

- The researchers used ComPAIR, an open source tool to run an online experiment in which participants were asked, for each pair of learner texts, to decide which of the two texts was written by a more advanced speaker.
- The researchers collected 318 comparisons by 43 participants (called judges in the study) based on 50 learner texts, all written on the same topic.
- Unlike in traditional crowdsourcing tasks, the researchers did not recruit the general public but solicited participants via professional mailing lists. They also asked participants to fill in a short questionnaire about their L1, proficiency in English, and previous experience and training in the evaluation of learners' writing.

What the researchers found

- A crowd of language specialists (university professors, PhD students, MA students, etc.) was able to assess learner texts with high reliability in an adaptive comparative judgment task.
- The final ranking generated by the crowdsourced comparative judgment task exhibited a strong positive association with the scores that resulted from professional rubric-based language assessment of these texts (as low, medium, and high writing proficiency).
- No effect of language competence or language assessment experience was found, but there seemed to be a difference between the decisions made by judges who had received formal language assessment training and those who had not.

Things to consider

- In this article about a new research method, the researchers illustrated how the technique of adaptive comparative judgment can be used to assess the proficiency of learner writing. This was done, however, on the basis of learner texts that shared a number of characteristics (same topic, similar length) and represented the full proficiency spectrum. Avenues for future investigation include replicating the experiment with a more diverse group of texts written by learners with similar language competence.
- The technique may also serve to elicit human comparative judgments of other constructs (e.g., complexity, accuracy, fluency, pronunciation) that have proved difficult to define and operationalize.

Materials, data, open access article: Data are publicly available on IRIS (<https://www.iris-database.org>) and OSF (<https://osf.io/89vm4/>).

How to cite this summary: Paquot, M. (2022). Adaptive comparative judgment is a viable option for proficiency rating in second language research. *OASIS Summary* of Paquot, Rubin, & Vandeweerd (2022) in *Language Learning*. <https://oasis-database.org>