

From Complexity Scores to Readable Texts: IREAD4SKILLS for Adult Literacy in Portuguese

Jorge Baptista^{1,2}, Eugénio Ribeiro^{1,3}, Nuno Mamede¹, David Antunes¹, Raquel Amaro⁴

¹ INESC-ID Lisboa, Portugal

{jorge.baptista,eugenio.ribeiro,nuno.mamede,david.f.l.antunes}@inesc-id.pt

² Faculdade de Ciências Humanas e Sociais, Universidade do Algarve, Portugal

³ Instituto Universitário de Lisboa (ISCTE-IUL), Portugal

⁴ CLUNL, Universidade NOVA de Lisboa, Portugal

raquelamaro@fcsn.unl.pt

Abstract

Adult Learning (AL) programmes need short, trustworthy texts that match learners’ reading abilities, but educators rarely have time, tools, or evidence-based guidelines to select and adapt materials consistently. We present a live demo of IREAD4SKILLS for European Portuguese: a web-based system that (i) estimates readability/complexity for AL-oriented levels aligned with CEFR, (ii) highlights where complexity concentrates (lexical, grammatical, semantic), and (iii) supports rewriting by offering actionable, level-aware suggestions and curated lexical resources. The demo emphasises transparency and “trainer-first” workflows: users see *why* a text is complex and *how* to revise it without losing meaning.

1 Motivation

Readability assessment has moved well beyond classic formulas, toward feature-rich and neural approaches that can support practical interventions in educational settings (Collins-Thompson, 2014; Crossley et al., 2019; Lee et al., 2021). In Adult Learning, the target is not “school-grade” readability but *functional* comprehension and task-oriented reading under time constraints, often for learners with limited literacy backgrounds (OECD, 2024, 2023, 2013; Steeds, 2001). For European Portuguese, recent work has shown the feasibility of automatic readability modelling with supervised learning and a careful feature design (Ribeiro et al., 2024a,b), while related efforts address generalisation and feature transfer across Portuguese settings (Akef et al., 2024).

Our contribution in this demo paper is the *interactive bridge* from a readability score to concrete, level-driven rewriting: users can (i) paste/upload a text (or even a mobile phone picture of it, to be OCR processed), (ii) obtain an overall level predic-

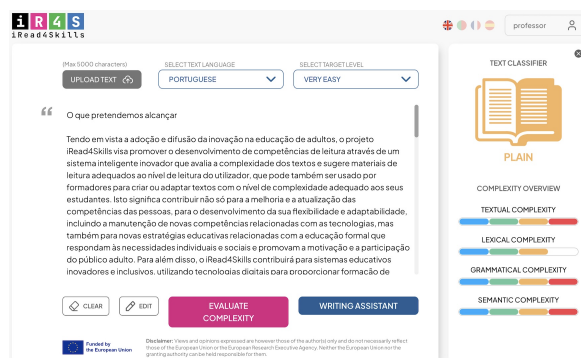


Figure 1: Classifier view: overall level prediction with a compact breakdown by complexity dimension (Portuguese example).

tion with an interpretable breakdown, and (iii) iteratively revise the same text guided by explanations grounded on project research and resources (Amaro et al., 2025; Monteiro et al., 2025; Blanco Escoda et al., 2023; Amaro et al., 2024). The intended audience includes low-literacy adults, AL trainers, curriculum designers, librarians, publishers, and researchers interested in Portuguese readability and explainable educational NLP.

2 IREAD4SKILLS System in one minute

IREAD4SKILLS is an open-access platform that supports multilingual complexity assessment of texts; here we focus on the Portuguese pipeline and AL use cases. A typical interaction is: (i) select language and desired target level (e.g., Portuguese, *Very Easy*), (ii) submit a text (up to a short passage), (iii) inspect the predicted level and the four-dimensional overview (textual, lexical, grammatical, semantic), and (iv) open the writing assistant to see sentence and token-level signals and access targeted resources.

Figure 1 illustrates the *Text Classifier* view, which returns the predicted level and a compact

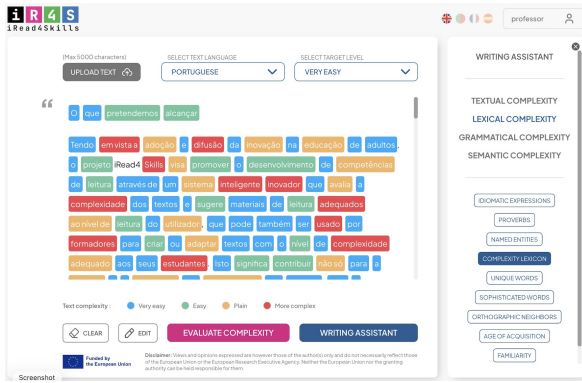


Figure 2: Writing assistant view: token-level highlighting and access to level-oriented resources (e.g., lexicons, unique/sophisticated words).

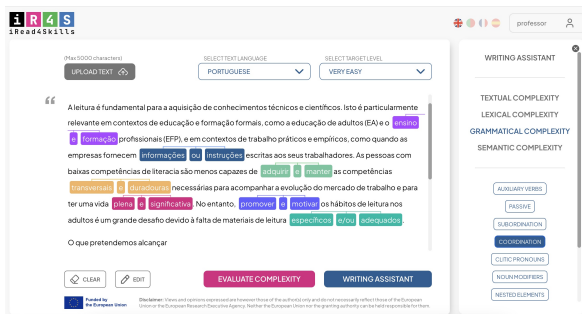


Figure 3: Writing assistant view: sequence-level highlighting coordinated elements.

complexity overview, across 4 yardsticks: Textual, Lexical, Grammatical, and Semantic Complexity. Figures 2 to Figure 4 show the *Writing Assistant* view, where tokens and sequences, respectively, are highlighted according to their contribution to complexity (e.g., unusually rare items, specialized vocabulary, long-distance syntactic dependencies, or semantic density).

3 Under the hood

The Portuguese pipeline follows a practical design aimed at AL settings: robust text processing and an interpretable feature layer, combined with modern modelling where adopted. At a high level, the system uses (i) curated corpora and level definitions tailored to iRead4Skills, including open-access datasets and lexicons by complexity level (Pintard et al., 2024b,a; Wilkens et al., 2024); (ii) feature sets grounded in readability research (e.g., lexical rarity/frequency, cohesion and discourse signals, and syntactic complexity proxies) (Chen and Meurers, 2018; McNamara et al., 2010; Crossley et al., 2017); and (iii) supervised readability models for European Portuguese aligned with CEFR-inspired

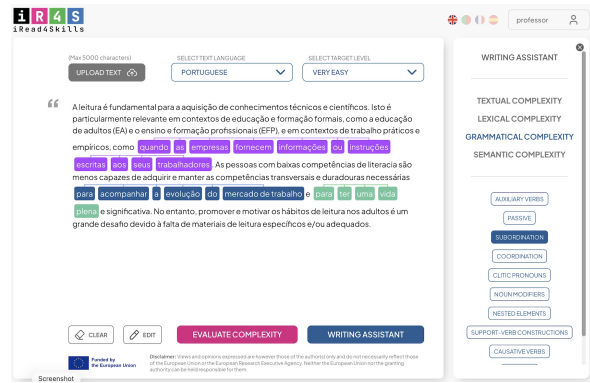


Figure 4: Writing assistant view: sequence-level highlighting with long-distance dependencies (subclauses).

targets (Council of Europe, 2020; Ribeiro et al., 2024a,b). This combination supports not only prediction but also explanation: the interface surfaces *what* is complex and offers *revision directions* consistent with the intended level.

Crucially, the demo is not presented as a black box. Instead, it operationalizes a “trainer-first” loop: (i) detect complexity hotspots, (ii) consult level-specific lexical resources and guidance, (iii) revise, and (iv) re-evaluate—a workflow aligned with educational NLP tools for complexity analysis (Chen and Meurers, 2016) and the broader readability literature (Collins-Thompson, 2014; Crossley et al., 2019).

4 Demo plan

The demo is designed for a fast, hands-on experience and for audience participation. In the session, we will run three short scripts with live interaction: (i) *cold-start diagnosis*—attendees paste a short Portuguese text (e.g., a leaflet, civic information, workplace safety note) and the system returns level + breakdown; and (ii) *guided rewriting*—we open the writing assistant and determine different constructs contributing to the text’s complexity. The original text can be rewritten and its complexity reassessed to meet the target user literacy level.

To maximise engagement, we will keep a “challenge bowl” of real-world micro-texts (one per domain: health, public services, work, training materials). Participants can vote for the most effective revision and we re-run the classifier to show before/after deltas. When feasible, we also show cross-language comparability by briefly switching to another supported language, reinforcing the platform’s multilingual scope.

5 Impact and availability

By aligning readability assessment with AL practice, iREAD4SKILLS supports: (i) faster selection of appropriate materials for low-literacy adult learners, (ii) consistent adaptation decisions backed by explainable signals, and (iii) resource creation and curation (lexicons and corpora) for the Portuguese community. The platform is openly accessible, and major project resources (datasets, lexicons, and level definitions) are distributed via project website¹ and Zenodo.²

Acknowledgments

Work supported by European Commission under Project iRead4Skills, HORIZON-CL2-2022-TRANSFORMATIONS-01-07, DOI: [10.3030/101094837](https://doi.org/10.3030/101094837), and by national funds through Fundação para a Ciência e a Tecnologia, I.P. (FCT) under projects UID/50021/2025 (DOI: <https://doi.org/10.54499/UID/50021/2025>), UID/PRR/50021/2025 (DOI: <https://doi.org/10.54499/UID/PRR/50021/2025>) and UID/03213/2025 (DOI: <https://doi.org/10.54499/UID/03213/2025>).

References

- S. Akef, A. Mendes, D. Meurers, and P. Rebuschat. 2024. Investigating the Generalizability of Portuguese Readability Assessment Models Trained Using Linguistic Complexity Features. In *PROPOR 2024*, pages 332–341.
- R. Amaro, S. Correia, R. Monteiro, A. Pintard, M. Mourinho, and S. Barbosa. 2025. Cadre de référence pour la complexité textuelle destiné aux adultes peu alphabétisés : niveaux et descripteurs dans le cadre du projet iread4skill. *Langues & Parole(s)*, 10:57–119.
- Raquel Amaro, Ricardo Monteiro, Thomas François, and J. Nagant de Deuxchaisnes. 2024. *iread4skills - data set 2: Annotated corpora report*.
- Xavier Blanco Escoda, Raquel Amaro, Thomas François, and Marcos Garcia. 2023. *iread4skills - baselines for complexity lexicons definition*.
- X. Chen and D. Meurers. 2016. CTAP: A Web-based Tool Supporting Automatic Complexity Analysis. In *CLALC*, pages 113–119.
- X. Chen and D. Meurers. 2018. Word Frequency and Readability: Predicting the Text-level Readability with a Lexical-level Attribute. *Journal of Research in Reading*, 41(3):486–510.
- K. Collins-Thompson. 2014. Computational assessment of text readability: A survey of current and future research. *International Journal of Applied Linguistics*, 165(2):97–135.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion Volume*. Council of Europe Publishing, Strasbourg.
- S. A. Crossley, S. Skalicky, and M. Dascalu. 2019. Moving beyond classic readability formulas: New methods and new models. *Journal of Research in Reading*, 42(3-4):541–561.
- S. A. Crossley, S. Skalicky, M. Dascalu, D. S. McNamara, and K. Kyle. 2017. Predicting text comprehension processing and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, 54(5-6):340–359.
- B. W. Lee, Y. S. Jang, and J. Lee. 2021. Pushing on text readability assessment: A transformer meets handcrafted linguistic features. In *EMNLP 2021*, pages 10669–10686.
- D. S. McNamara, M. M. Louwerse, P. M. McCarthy, and A. C. Graesser. 2010. Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4):292–330.
- R. Monteiro, S. Correia, R. Amaro, M. Moutinho, S. Barbosa, and M. L. Reis. 2025. *Níveis e descritores de complexidade textual para adultos de baixa literacia: um referencial do projeto iread4skills*. *Revista da Associação Portuguesa de Linguística*, 1(13):193–222.
- OECD. 2013. *The Survey of Adult Skills Reader's Companion*. OECD Publishing, Paris.
- OECD. 2023. *PISA 2022 Results (Volume I): The State of Learning and Equity in Education, PISA*. Technical report, OECD Publishing, Paris.
- OECD. 2024. *Do Adults Have the Skills They Need to Thrive in a Changing World?: Survey of Adult Skills 2023*. Technical report, OECD Publishing, Paris.
- A. Pintard, T. François, J. Nagant de Deuxchaisnes, S. Barbosa, M. L. Reis, M. Moutinho, R. Monteiro, R. Amaro, S. Correia, S. Rodríguez Rey, K. Mu, M. Garcia González, A. Bernárdez Braña, and X. Blanco Escoda. 2024a. *iread4skills dataset 2: annotated corpora by level of complexity for fr, pt and sp*.
- A. Pintard, Thomas François, J. Nagant de Deuxchaisnes, Sílvia Barbosa, M. L. Reis, Michel Moutinho, Ricardo Monteiro, Raquel Amaro, Sara Correia, S. Rodríguez Rey, Marcos Garcia González, K. Mu, and Xavier Blanco Escoda. 2024b. *iread4skills dataset 1: corpora by complexity level for fr, pt and sp*.
- E. Ribeiro, N. Mamede, and J. Baptista. 2024a. Automatic Text Readability Assessment in European Portuguese. In *PROPOR 2024*, pages 97–107.
- E. Ribeiro, N. Mamede, and J. Baptista. 2024b. Text Readability Assessment in European Portuguese: A Comparison of Classification and Regression Approaches. In *PROPOR 2024*, pages 551–557.
- Andrew Steeds. 2001. *Adult Literacy Core Curriculum Including Spoken Communication*. ERIC/The Basic Skills Agency, Commonwealth House, London.
- R. Wilkens, A. Pintard, T. François, S. Barbosa, M. L. Reis, R. Amaro, E. Ribeiro, N. Mamede, J. Baptista, X. Blanco, A. Catena, R. Gauchola, and K. Mu. 2024. *iread4skills - basic lexicons per complexity level*.

¹<https://iread4skills.com/tools-resources>

²<https://zenodo.org/communities/iread4skills/>