Comparison of Tools for Evaluating Text Progression

This document summarises a conversation about using four different tools to evaluate the CEFR levels of a sequence of texts designed to help learners extend their language repertoire for speaking tasks. The tools compared were ChatGPT, GSE Text Analyzer, Text Inspector CEFR, and Write & Improve CEFR.

Observations from the Graph

The graph provided shows the progression of texts through different levels of complexity, evaluated by the tools. Key observations include:
- ChatGPT generally assigns lower CEFR levels, suitable for spoken-style texts.
- GSE Text Analyzer provides moderate estimates but lacks granularity for advanced levels.
- Text Inspector CEFR overestimates complexity, particularly for simpler texts.
- Write & Improve CEFR tracks progression realistically, making it the most reliable for balanced evaluation.

Evaluation of the Approach

The aim of the text progression was to provide learners with suggestions for extending their repertoire so that they could produce more sophisticated accounts when repeating tasks or performing similar tasks.

What worked:
- The progression effectively scaffolds learning, starting with a corrected baseline and moving to advanced versions.
- Write & Improve supports the progression by reflecting noticeable growth.
- New features in 'One level up' and 'Two levels up' texts provide learners with manageable challenges.

Challenges:
- Text Inspector's overestimation could mislead learners by assigning overly high levels.
- ChatGPT and GSE Text Analyzer might not fully recognise progress in advanced texts, which could be discouraging.

Suggestions for Future Iterations

1. Ensure progression includes structures and vocabulary learners can realistically integrate.
2. Use tools like Write & Improve to provide concrete feedback on improvements and areas for development.
3. Complement automated evaluations with manual analysis for spoken discourse features.
4. Encourage repetition of tasks, incorporating new language elements each time.

Conclusion

Write & Improve most closely mirrors the intended progression, making it the best choice for monitoring language complexity growth. ChatGPT serves as a good baseline for spoken-level appropriateness, ensuring realistic targets.