

REPO TEST 01 / TESTED 2026-10-03

Ponytail wrote less code. Our small test did not prove it was better.

THE QUICK TAKE

Ponytail reduced added lines from 13 to 9 in one small coding task. Both setups reused the existing helpers and passed the same three tests. The Ponytail run had a higher estimated list-price cost, so this is an interesting result, not evidence of a general win.

The setup

Version: 4.10.3 / c982cd4 Scope: One small CSV task, one run per setup, Claude Opus 5.5.

Measure	Without Ponytail	With Ponytail
Added / removed lines	13 / 2	9 / 2
Tests passing	3 of 3	3 of 3
Wall time	14.8 seconds	13.1 seconds
Output tokens	1,479	1,287
Estimated list-price cost	\$0.1089	\$0.1247

Verdict: Interesting, still inconclusive

One run per setup; no repeated benchmark. Both versions passed 3 tests. The plugin loaded, but its broader savings claims were not reproduced.

KEEP THE METHOD / MAKE IT YOUR OWN

Plan your own paired repo test

A reusable planning prompt. Replace bracketed details with your own. This prompt is not an executed test or a guarantee of results.

Review [REPOSITORY] and its claims before running it. Use a disposable project with synthetic data. Pin the repo commit and the model version.

Choose one realistic task where unnecessary code would be tempting. Write the acceptance checks before starting. Run the same task both with and without the tool, with the same instructions, model, permissions and starting files.

Keep each run isolated. Repeat enough times to show variation; report every run, not just the best one. Record setup work, final diffs, passing and failing checks, elapsed time and token usage.

Distinguish estimated list-price cost from money actually paid. Check the resulting work independently. Explain what the test does not prove.

Do not install globally, use real credentials or customer data, spend money, or publish anything. Stop and explain any missing prerequisites.

Keep with your result

Record the exact version, date, starting conditions, acceptance checks, measured outputs, and failures. State what was not tested. Keep private logs and credentials out of public downloads.

Source and full report

<https://github.com/DietrichGebert/Ponytail>

<https://www.everyoneneedsasamwise.com/articles/ponytail-repo-test>

Measurements are from ENAS local evidence recorded on October 3, 2026. The article contains the fuller method and caveats. The downloadable prompt was prepared with this report.