

```
# Data Report — v2 run- Generated: 2025-09-07 03:23- Pairs (n): 60- Provider / Model /
Seed: ollama / gemma:7b / 42## Aggregate means (per-id)| metric | mean_base |
mean_instr | Δ | n(items) || --- | --- | --- | --- | --- || BLEU | 19.559 | 25.190 | 5.631 | 60 ||
ROUGE-L | 0.284 | 0.287 | 0.004 | 60 || chrF | 36.118 | 33.485 | -2.633 | 60 |## Paired
statistics| metric | n | mean_base | mean_instr | Δ | Δ% | d | CI95_low | CI95_high | p | q || --
- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- || BLEU | 60 | 19.559 | 25.190 | 5.631 |
28.792 | 0.617 | 3.456 | 8.054 | 0.000 | 0.000 || ROUGE-L | 60 | 0.284 | 0.287 | 0.004 | 1.289 |
0.038 | -0.021 | 0.027 | 0.621 | 0.831 || chrF | 60 | 36.118 | 33.485 | -2.633 | -7.291 | -0.187 |
-6.242 | 0.781 | 0.831 | 0.831 |### Latency (preview)| file | n | latency_ms_mean |
latency_ms_p50 | latency_ms_p95 | latency_ms_min | latency_ms_max | tokens_mean || --- | -
-- | --- | --- | --- | --- | --- | --- || general.jsonl | 60 | 10167.4 | 10137.0 | 19436 | 3307.0 |
24167.0 | 0.0 || general_pass.jsonl | 60 | 10706.6 | 10992.0 | 20507 | 3532.0 | 22106.0 | 0.0 ||
instructed.jsonl | 60 | 8100.9 | 7768.0 | 13012 | 2994.0 | 15490.0 | 0.0 || instructed_pass.jsonl
| 60 | 9081.9 | 8582.5 | 14033 | 3356.0 | 19383.0 | 0.0 |### Compliance delta (pass)- same:
60, diff: 0, only_base: 0, only_cvd: 0
```

---

## \*\*Notes\*\*

- Aggregate means above are computed from per-id files (`bleu\_sacre.json`, `chrF.json`, `rouge.json`) to avoid folder contamination.- Paired statistics are taken from `results/quantitative/stats\_summary.v2.csv` (bootstrap CI, Wilcoxon, BH-FDR).- For detailed cases, open `figs/error\_board.html` (Top/Bottom ΔBLEU examples).