

TECHNICAL FINDINGS REPORT

ABAD Adaptive ZD30 v4.4-v4.5

Language-Origin Screen and G2-S1 Middle English Mapping

Date	15 August 2026
Structural unit	Four frozen v4.3 candidate contribution blocks
Screen result	Middle English ranked first in all four blocks
Mapping decision	FAIL — G2-S1 controlled Middle English
Boundary	No language identification or decryption

Executive finding

Outcome: The low-flexibility screen consistently ranked Middle English nearest, but the full controlled mapping failed. G2-S1 barely cleared lexical coverage while failing structural coherence, phase recurrence, page-null, and evidence-balance gates.

This is exactly why the cascade separates candidate ranking from validation. The structural folio segmentation remains intact. The current Middle English substitution + controlled-variation mechanism is rejected for the cleanest writer-like block.

Gate	Observed	Requirement	Result
Selection-aware p	0.00498	< 0.05	PASS
Lexical coverage	41.07%	≥ 40%	PASS
Structural unit z	-0.809	> 0	FAIL
Phase recurrence z	-3.327	> 0	FAIL
Maximum page null	81.40%	≤ 60%	FAIL
Largest evidence share	50.32%	≤ 50%	FAIL

Bottom line: The low p-value does not rescue the mapping. It mainly reflects unusually high dictionary collision coverage; the mapped labels behave worse than nulls on the structural and phase units.

1. v4.4 language-origin screen

Block	Folios	Unique types	Nearest reference	Distance	Train/held rank
G1-S1	52	1,689	Middle English	10.98	1 / 1
G1-S2	30	1,275	Middle English	11.33	1 / 1
G2-S1	84	1,062	Middle English	13.90	1 / 1
G2-S2	57	1,363	Middle English	12.29	1 / 1

All four blocks produced the same order: Middle English, Old English, modern French control, modern Italian control, incomplete medieval Latin. The repeated ranking is stable across the fixed folio halves.

Critical limitation: The screen is substitution-invariant and dictionary-independent, but the references are unequal. Middle English has a large historical inventory; Old English is poetry-based; Latin is severely incomplete; French and Italian are modern controls. Therefore the ranking prioritizes tests—it does not identify a language.

2. What v4.4 does not support

- It does not support four different base languages; every block ranked the same way.
- It does not rule out Latin, French, Italian, German, or another language because comparable historical corpora were unavailable.
- It does not turn word-shape similarity into readable plaintext.
- It does not validate care/caring or any earlier proposed anchor.

3. v4.5 G2-S1 full mapping test

G2-S1 was chosen before mapping because it is the cleanest candidate contribution: 84 folios, 81 assigned to Hand 1, with 99.88% structural assignment stability. The exact v1.13 Middle English lexicon, controlled scribal variations, abbreviation inventory, complexity costs, five scoring units, and 200 complete mapping-search nulls were retained.

Scoring unit	z-score	Interpretation
Morphological label → next class	-0.809	Below matched null
Twelve-lock phase recurrence	-3.327	Strongly below matched null
Semantic label → illustration	-0.098	Near null
Held-out stem/suffix productivity	+0.560	Weak positive
Dictionary coverage	+4.855	Dominant positive unit

Why it failed: Coverage reached 41.07%, but semantic coverage was only 0.44%. No verbose spans survived. At least one page was 81.4% unresolved, and the two primary structural units were negative. The mapping creates dictionary forms without creating a coherent language-like sequence.

The complexity-penalized composite exceeded every null composite, producing $p = 0.00498$. Nevertheless, the preregistered protocol requires every mandatory gate. This protects against declaring success from a single flexible coverage channel.

4. Scientific interpretation

Still supported	Rejected or unsupported
Two primary structural modes and four residual contribution blocks	Pure Middle English under the frozen v1.13 mapping on G2-S1
G2-S1 is a coherent nearly Hand-1 corpus	A coherent English grammatical/semantic reading
Blockwise testing is methodologically superior to pooling	The current screen as proof of English origin
Further languages can be tested on unchanged blocks	A multilingual-scribe conclusion from present data

5. Where this leaves the Voynich project

Honest status: We have made defensible progress in segmentation, not decryption. The present result moves us farther from an unsupported English claim and closer to a valid experiment on language origin.

The strongest new object is the frozen contributor atlas. Any future mapping must explain why four structurally coherent blocks exist and must outperform the failed G2-S1 English baseline on untouched folios.

6. Next logical work

- Acquire comparable 14th-15th-century corpora for medieval Latin, Middle High German, Old/Middle French, and Northern Italian before ranking them.
- Repeat the same substitution-invariant screen with matched corpus sizes, genres, and frequency information.
- Test a language only when its historical corpus ranks consistently and its mapping rules are frozen in advance.
- Compare one shared language against block-specific languages with an explicit penalty for every additional language.
- Do not expand homophony, nulls, or abbreviation freedom merely to repair the failed English run.

Recommended next protocol: A corpus-quality upgrade comes before another mapping search. The present Latin/French/Italian references are not adequate for a fair multilingual decision.

Integrity

Receipt	Hash
v4.4 language screen	ccaafc7a4cc906cd916534fbd2aaf57b5f91b5a67819f3ee7ece43d0d13410e2
v4.5 G2-S1 mapping	414424abbf49f73076331d6c6915fd59b04b0ad8b6f6741f6784c1eadc88ea6e