Table 1: Summary of RQ3 build time tests

Metric	Wilcoxon test	Cliff's Estimate test
Average Hierarchy Height	0.7658	0.04861111 (negligible)
High-level Structuring	0.01626	-0.4027778 (medium)
Class Structuring	0.05782	-0.3211806 (small)
Operation Structuring	0.736	-0.05902778 (negligible)
Coupling Intensity	0.3104	-0.1736111 (small)
Intrinsic Operation Complexity	0.3368	0.1631944 (small)
Coupling Dispersion	0.02451	0.3802083 (medium)

Table 2: Summary of RQ3 build fix time tests

Metric	Wilcoxon test	Cliff's Estimate test
Average Hierarchy Height	0.001594	-0.5256917 (large)
High-level Structuring	0.5702	-0.1007905 (negligible)
Class Structuring	0.04099	-0.3577075 (medium)
Operation Structuring	0.4202	0.1422925 (negligible)
Coupling Intensity	0.166	0.243083 (small)
Intrinsic Operation Complexity	0.2024	-0.2233202 (small)
Coupling Dispersion	0.3341	-0.1699605 (small)

Table 3: Summary of RQ3 test coverage tests

Metric	Wilcoxon test	Cliff's Estimate test
Average Hierarchy Height	0.1817	0.2152778 (small)
High-level Structuring	0.2357	-0.2013889 (small)
Class Structuring	0.3481	0.1597222 (small)
Operation Structuring	0.004501	-0.4722222 (medium)
Coupling Intensity	6.617e-06	-0.7604167 (large)
Intrinsic Operation Complexity	0.0009713	0.5555556 (large)
Coupling Dispersion	0.001234	0.5451389 (large)

Table 4: Summary of RQ3 CI activity tests

Metric	Wilcoxon test	Cliff's Estimate test
Average Hierarchy Height	0.4654	0.1197917 (negligible)
High-level Structuring	0.2524	-0.1944444 (small)
Class Structuring	0.1904	-0.2222222 (small)
Operation Structuring	0.6134	0.08680556 (negligible)
Coupling Intensity	0.2085	0.2135417 (small)
Intrinsic Operation Complexity	0.2643	-0.1892361 (small)
Coupling Dispersion	0.4209	-0.1371528 (negligible)