

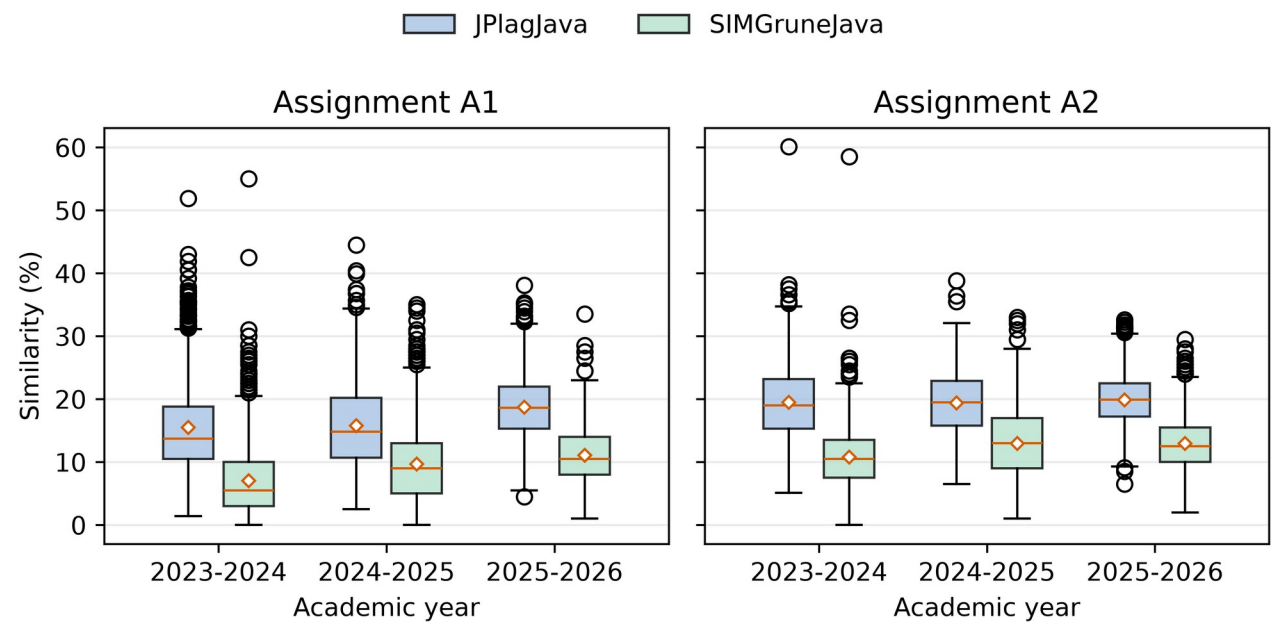
Supplementary Materials for Rethinking Similarity Scores in Programming Education: A Three-Year Study Across Assignments and Detection Tools

These Supplementary Materials provide additional analyses supporting the results reported in the main manuscript. They include the original pairwise similarity-score distributions, detailed submission-level descriptive statistics, a sensitivity analysis excluding students who appeared in more than one academic year, empirical cumulative distribution functions (ECDFs), threshold sensitivity analyses, an analysis of the 100 highest pairwise similarity scores, and a matched cross-tool rank-association analysis.

Contents:

- Figure S1. Pairwise similarity score distributions across academic years, assignments, and tools.
- Table S1. Complete submission-level descriptive statistics.
- Table S2. Complete submission-level descriptive statistics excluding repeated students (sensitivity analysis).
- Figure S2. Submission-level mean similarity across academic years, assignments, and tools.
- Figure S3. Submission-level maximum similarity across academic years, assignments, and tools.
- Figures S4 - S7. Empirical cumulative distribution functions (ECDFs) by assignment and detection tool.
- Tables S3 - S4. Threshold sensitivity analysis of pairwise similarity scores for Assignments A1 and A2.
- Table S5. Top-100 pairwise similarity score analysis.
- Table S6. Cross-tool rank association between JPlagJava and SIMGruneJava.
- Figure S8. Association between JPlagJava and SIMGruneJava pairwise similarity scores across assignments and academic years.

Figure S1. Pairwise similarity score distributions across academic years, assignments, and tools. These distributions correspond to the pairwise descriptive statistics reported in Tables 1 and 2 of the main manuscript.



Note: Boxes show the interquartile range, horizontal lines indicate medians, and diamonds indicate means. Whiskers extend to the most extreme observations within 1.5 interquartile ranges of the box; circles indicate observations beyond the whiskers.

Table S1. Complete submission-level descriptive statistics.

Tool	Assignment	Year	Metric	n	mean	median	Q1	Q3	SD	min	max
JPlag	A1	2023/2024	maximum	52	30.38	30.90	23.95	35.88	9.07	13.80	51.90
JPlag	A1	2024/2025	maximum	59	29.27	30.00	25.80	33.50	7.50	12.20	44.50
JPlag	A1	2025/2026	maximum	55	28.82	28.30	25.70	32.50	4.76	16.90	38.10
JPlag	A2	2023/2024	maximum	52	30.60	30.10	27.40	32.75	7.53	15.70	60.10
JPlag	A2	2024/2025	maximum	56	28.10	28.30	24.68	31.43	5.08	15.60	38.80
JPlag	A2	2025/2026	maximum	54	27.68	28.15	25.63	30.10	3.25	15.90	32.60
SIM	A1	2023/2024	maximum	52	23.30	22.50	15.50	28.50	9.84	9.00	55.00
SIM	A1	2024/2025	maximum	59	22.81	23.00	18.00	27.50	7.45	4.00	35.00
SIM	A1	2025/2026	maximum	55	19.92	19.50	18.00	22.00	4.65	9.50	33.50
SIM	A2	2023/2024	maximum	52	22.67	20.75	18.50	24.00	8.67	8.00	58.50
SIM	A2	2024/2025	maximum	56	22.30	22.50	19.00	25.50	6.24	6.50	33.00
SIM	A2	2025/2026	maximum	54	21.33	21.50	19.00	24.38	4.10	12.50	29.50
JPlag	A1	2023/2024	mean	52	15.48	15.33	13.13	18.17	3.30	6.57	21.21
JPlag	A1	2024/2025	mean	59	15.78	16.82	14.38	17.73	3.26	7.69	20.36
JPlag	A1	2025/2026	mean	55	18.73	19.09	16.92	20.34	2.62	11.97	23.96
JPlag	A2	2023/2024	mean	52	19.48	20.26	17.41	21.77	3.13	10.84	24.23
JPlag	A2	2024/2025	mean	56	19.40	20.13	17.54	21.32	2.83	11.68	23.47
JPlag	A2	2025/2026	mean	54	19.89	20.06	18.90	21.10	2.17	11.87	25.36
SIM	A1	2023/2024	mean	52	7.01	6.94	5.65	8.52	2.13	2.15	12.68
SIM	A1	2024/2025	mean	59	9.65	10.21	7.74	11.59	3.08	1.33	15.39
SIM	A1	2025/2026	mean	55	11.05	11.08	9.64	12.50	2.34	4.67	14.69
SIM	A2	2023/2024	mean	52	10.79	10.95	9.98	11.92	2.28	3.20	15.67
SIM	A2	2024/2025	mean	56	12.95	13.82	10.80	15.18	3.34	3.77	18.24
SIM	A2	2025/2026	mean	54	12.96	13.19	11.75	14.02	2.14	6.92	18.13
JPlag	A1	2023/2024	median	52	14.51	14.50	12.30	16.40	2.94	6.10	20.70
JPlag	A1	2024/2025	median	59	15.20	15.90	13.98	17.20	3.18	7.70	20.95
JPlag	A1	2025/2026	median	55	18.67	18.75	17.25	20.58	2.74	11.80	24.05
JPlag	A2	2023/2024	median	52	19.47	20.10	17.45	21.80	3.22	11.10	24.30
JPlag	A2	2024/2025	median	56	19.60	20.00	17.53	21.73	3.05	11.70	24.50
JPlag	A2	2025/2026	median	54	19.88	20.05	18.75	21.28	2.22	11.70	25.20
SIM	A1	2023/2024	median	52	6.02	5.75	4.50	7.63	2.30	2.00	11.50
SIM	A1	2024/2025	median	59	9.27	9.75	7.50	11.50	3.14	1.00	15.50
SIM	A1	2025/2026	median	55	10.90	11.00	9.50	12.50	2.35	5.00	15.00
SIM	A2	2023/2024	median	52	10.62	10.75	9.50	11.63	2.48	3.00	16.00
SIM	A2	2024/2025	median	56	13.25	14.00	11.00	15.50	3.60	3.50	19.50
SIM	A2	2025/2026	median	54	12.80	13.00	11.50	14.00	2.13	7.00	18.50

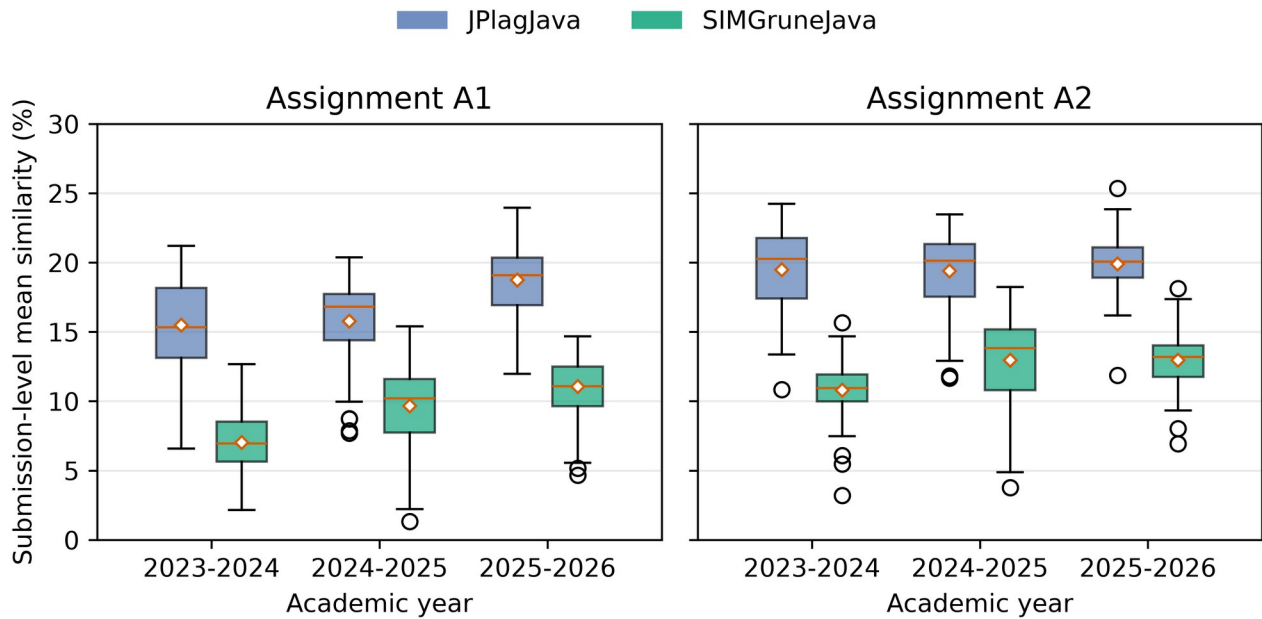
Note: JPlag = JPlagJava; SIM = SIMGruneJava. SIMGruneJava. Each row summarises the distribution of per-submission mean, median, or maximum similarity.

Table S2. Complete submission-level descriptive statistics excluding repeated students (sensitivity analysis)

Tool	Assignment	Year	Metric	n	mean	median	Q1	Q3	SD	min	max
JPlag	A1	2023/2024	maximum	48	28.59	28.95	21.83	35.18	7.50	13.80	41.90
JPlag	A1	2024/2025	maximum	51	29.21	30.10	25.80	33.40	7.57	12.20	44.50
JPlag	A1	2025/2026	maximum	51	28.47	28.30	25.65	31.85	4.72	16.90	38.10
JPlag	A2	2023/2024	maximum	47	29.58	29.90	27.30	32.55	4.90	15.70	38.20
JPlag	A2	2024/2025	maximum	49	27.45	28.20	24.70	30.00	4.53	15.60	36.40
JPlag	A2	2025/2026	maximum	52	27.56	27.90	25.88	29.50	3.18	15.90	32.60
SIM	A1	2023/2024	maximum	48	20.50	20.75	15.50	25.50	6.02	9.00	30.00
SIM	A1	2024/2025	maximum	51	21.62	22.00	16.50	26.50	7.12	4.00	35.00
SIM	A1	2025/2026	maximum	51	20.04	19.50	18.00	22.25	4.78	9.50	33.50
SIM	A2	2023/2024	maximum	47	20.49	20.50	18.50	24.00	4.33	8.00	26.50
SIM	A2	2024/2025	maximum	49	21.57	22.00	18.00	24.50	6.13	6.50	32.50
SIM	A2	2025/2026	maximum	52	21.32	21.25	18.75	24.00	4.01	12.50	29.50
JPlag	A1	2023/2024	mean	48	15.16	15.12	12.74	17.28	3.14	6.59	21.17
JPlag	A1	2024/2025	mean	51	15.84	16.72	14.61	17.70	3.19	7.76	20.19
JPlag	A1	2025/2026	mean	51	18.66	19.32	16.87	20.71	2.76	11.90	23.92
JPlag	A2	2023/2024	mean	47	20.06	20.88	17.83	22.45	3.18	11.15	24.85
JPlag	A2	2024/2025	mean	49	19.57	20.19	18.08	21.55	2.78	11.64	23.67
JPlag	A2	2025/2026	mean	52	19.95	20.05	18.89	21.18	2.23	11.89	25.47
SIM	A1	2023/2024	mean	48	6.86	6.64	5.58	8.34	2.18	2.30	12.76
SIM	A1	2024/2025	mean	51	9.35	9.64	7.76	11.48	3.08	1.27	14.65
SIM	A1	2025/2026	mean	51	11.22	11.29	10.06	12.62	2.40	4.79	14.93
SIM	A2	2023/2024	mean	47	10.82	10.88	9.98	11.98	2.39	3.29	15.88
SIM	A2	2024/2025	mean	49	12.71	13.64	10.71	15.10	3.39	3.86	18.06
SIM	A2	2025/2026	mean	52	13.09	13.22	11.80	14.09	2.13	6.98	18.25
JPlag	A1	2023/2024	median	48	14.25	14.15	12.28	16.23	2.74	6.20	20.10
JPlag	A1	2024/2025	median	51	15.25	15.80	14.13	17.05	3.01	7.70	20.75
JPlag	A1	2025/2026	median	51	18.67	19.05	17.03	20.60	2.89	11.80	24.05
JPlag	A2	2023/2024	median	47	20.14	21.05	18.13	22.60	3.27	11.30	25.10
JPlag	A2	2024/2025	median	49	19.78	20.20	18.55	22.00	2.97	11.60	24.65
JPlag	A2	2025/2026	median	52	19.99	20.30	18.98	21.40	2.28	11.70	25.50
SIM	A1	2023/2024	median	48	6.03	5.50	4.88	7.50	2.21	2.00	11.50
SIM	A1	2024/2025	median	51	9.01	8.75	7.13	11.50	3.14	1.00	15.25
SIM	A1	2025/2026	median	51	11.13	11.25	9.63	12.63	2.44	5.00	15.25
SIM	A2	2023/2024	median	47	10.69	10.75	9.25	12.13	2.56	3.00	16.00
SIM	A2	2024/2025	median	49	13.07	13.75	11.00	15.50	3.66	4.00	19.25
SIM	A2	2025/2026	median	52	12.92	13.00	11.50	14.00	2.13	7.00	18.50

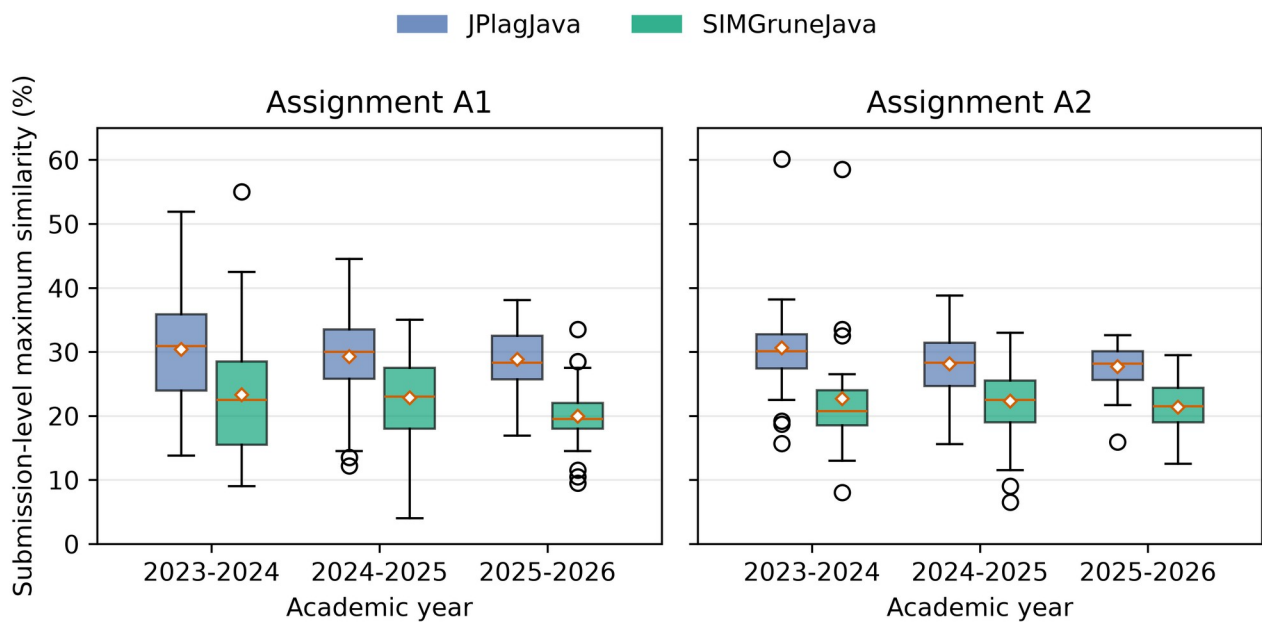
Note: JPlag = JPlagJava; SIM = SIMGruneJava. For each assignment, students appearing in multiple academic years and all their pairs were excluded before recalculation. Each row summarises the distribution of per-submission mean, median, or maximum similarity.

Figure S2. Submission-level mean similarity across academic years, assignments, and tools. For each submission, the mean similarity to all other submissions in the same assignment-year group was calculated. The figure shows the distribution of these per-submission mean values separately for Assignments A1 and A2, detection tools, and academic years.



Note: Boxes show the interquartile range, horizontal lines indicate medians, and diamonds indicate means. Whiskers extend to the most extreme observations within 1.5 interquartile ranges of the box; circles indicate observations beyond the whiskers.

Figure S3. Submission-level maximum similarity across academic years, assignments, and tools. For each submission, the maximum similarity to any other submission in the same assignment-year group was calculated. The figure shows the distribution of these per-submission maximum values separately for Assignments A1 and A2, detection tools, and academic years.



Note: Boxes show the interquartile range, horizontal lines indicate medians, and diamonds indicate means. Whiskers extend to the most extreme observations within 1.5 interquartile ranges of the box; circles indicate observations beyond the whiskers.

Figure S4. Empirical cumulative distribution functions for Assignment A1 using JPlagJava. The curves show the cumulative proportion of pairwise similarity scores at or below each similarity value for the three academic years.

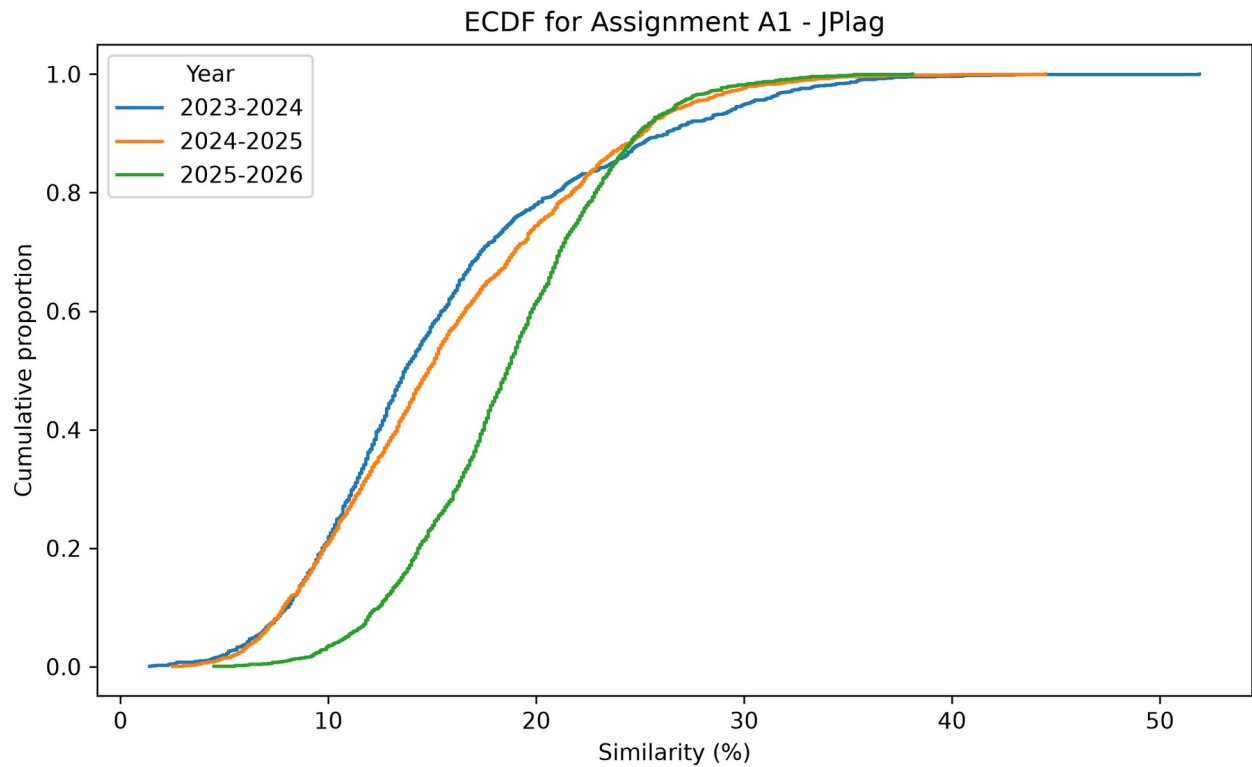


Figure S5. Empirical cumulative distribution functions for Assignment A1 using SIMGruneJava. The curves show the cumulative proportion of pairwise similarity scores at or below each similarity value for the three academic years.

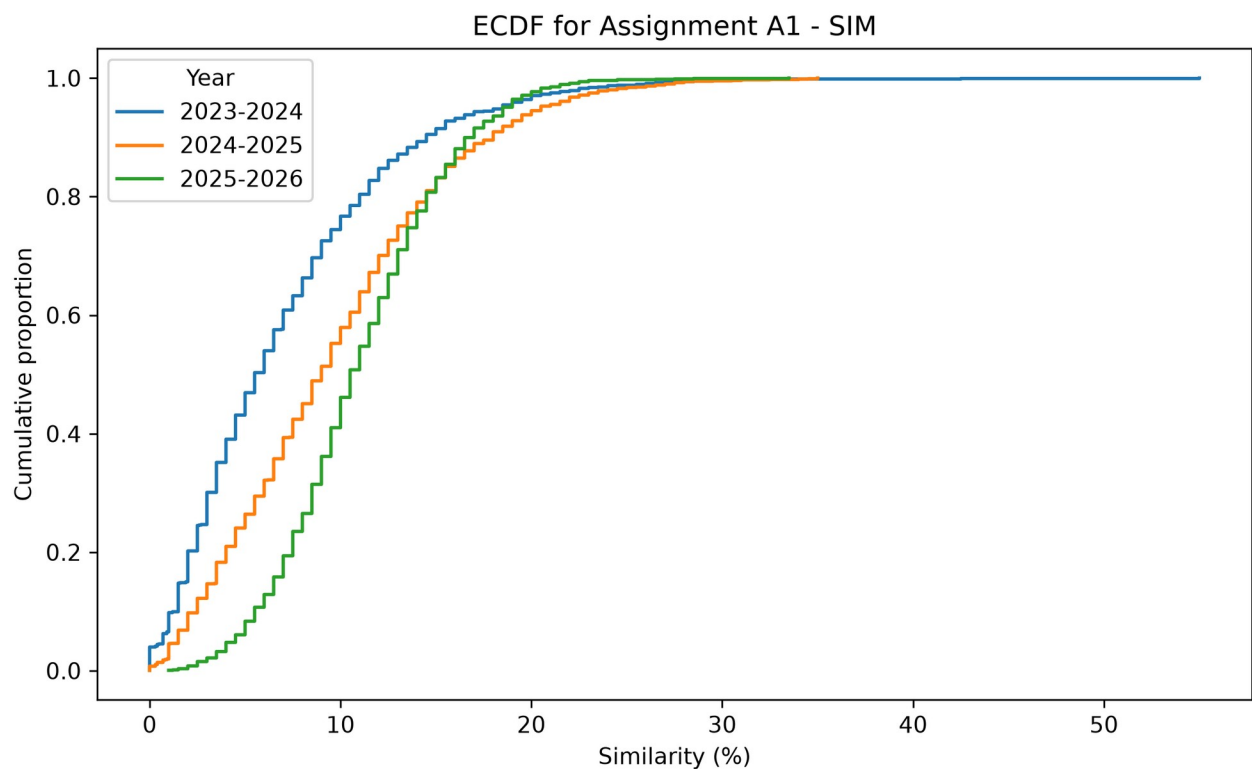


Figure S6. Empirical cumulative distribution functions for Assignment A2 using JPlagJava. The curves show the cumulative proportion of pairwise similarity scores at or below each similarity value for the three academic years.

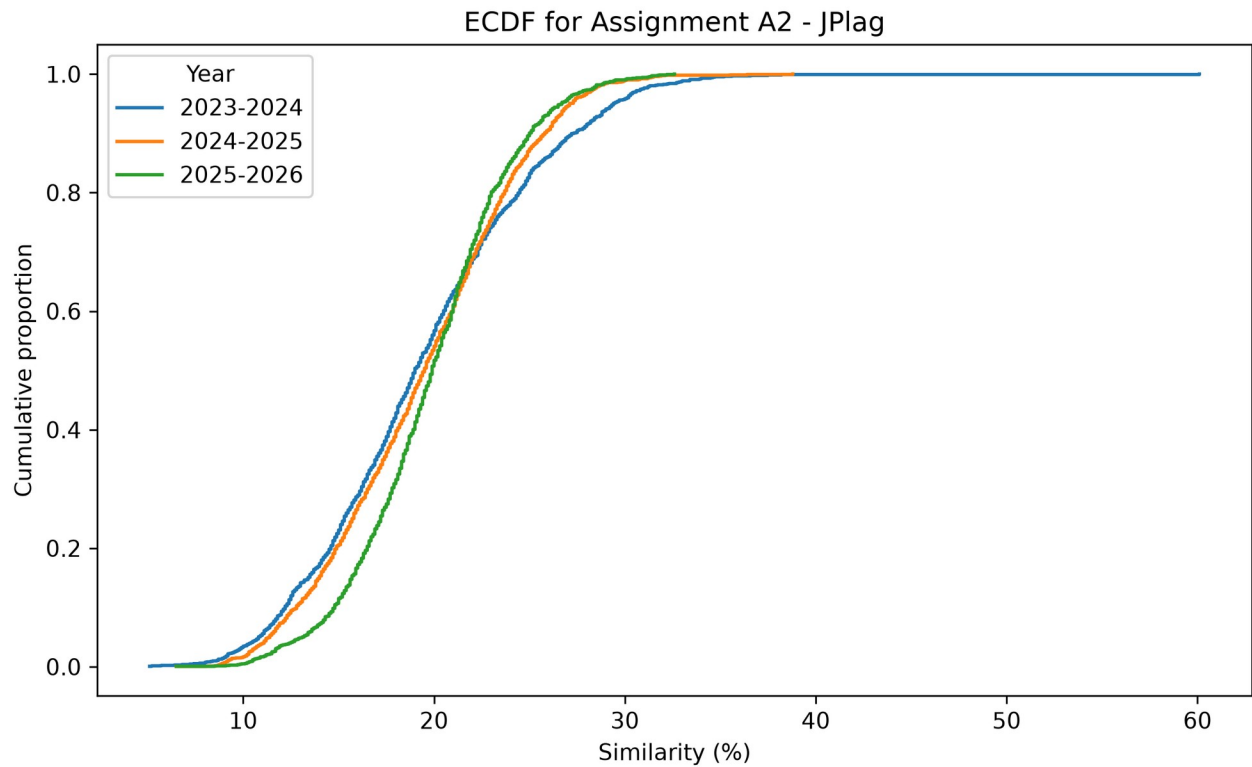


Figure S7. Empirical cumulative distribution functions for Assignment A2 using SIMGruneJava. The curves show the cumulative proportion of pairwise similarity scores at or below each similarity value for the three academic years.

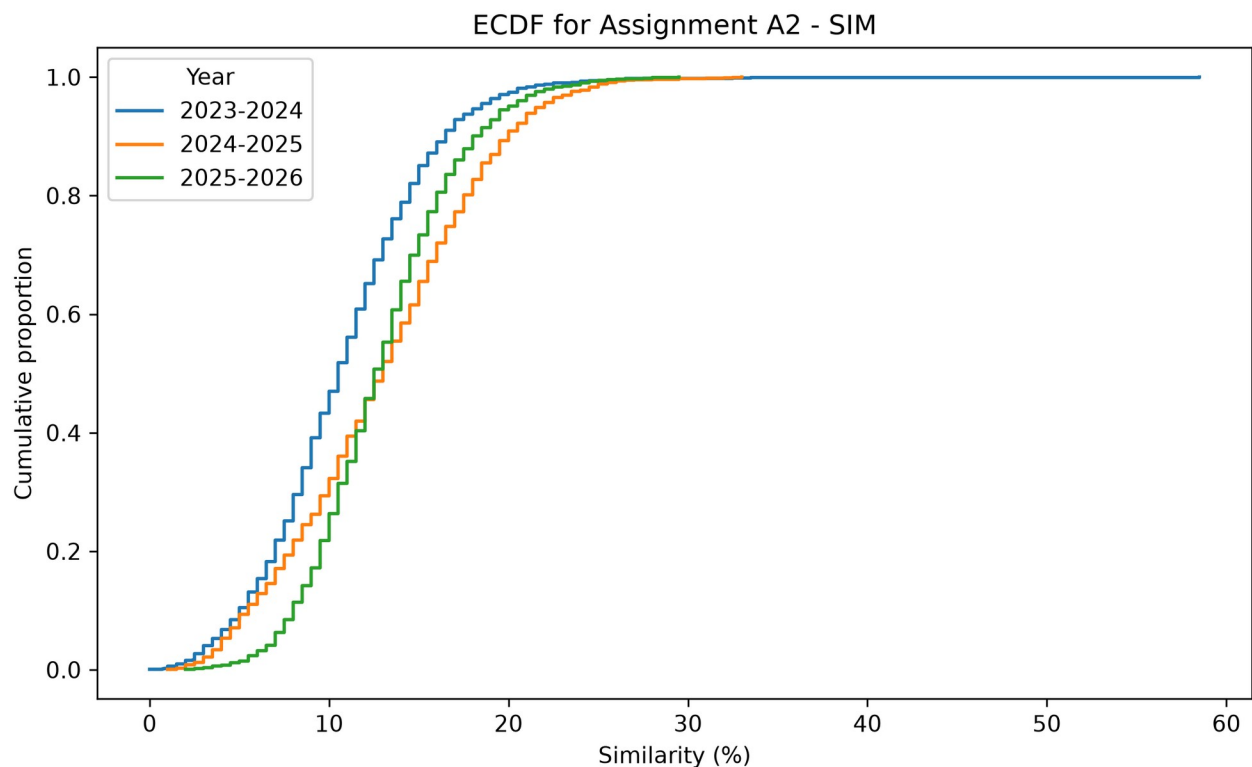


Table S3. Threshold sensitivity analysis of pairwise similarity scores for A1

Tool	Academic Year	≥15%	≥20%	≥25%	≥30%	≥35%	≥40%	≥45%	≥50%
JPlagJava	2023/2024	42.68	22.10	11.92	5.20	1.51	0.38	0.08	0.08
JPlagJava	2024/2025	49.33	25.66	10.46	2.40	0.47	0.12	0.00	0.00
JPlagJava	2025/2026	76.70	38.86	9.83	1.82	0.27	0.00	0.00	0.00
SIMGruneJava	2023/2024	9.50	3.62	1.21	0.30	0.15	0.15	0.08	0.08
SIMGruneJava	2024/2025	18.99	6.20	1.75	0.47	0.12	0.00	0.00	0.00
SIMGruneJava	2025/2026	19.26	2.90	0.27	0.07	0.00	0.00	0.00	0.00

Note: Values represent the percentage of all pairwise similarity scores at or above each threshold. Thresholds from 15% to 50% in five-percentage-point increments were examined to assess whether the observed patterns depend on the particular fixed thresholds used in the main analysis.

Table S4. Threshold sensitivity analysis of pairwise similarity scores for A2

Tool	Academic Year	≥15%	≥20%	≥25%	≥30%	≥35%	≥40%	≥45%	≥50%
JPlagJava	2023/2024	77.53	44.04	17.50	4.30	0.45	0.08	0.08	0.08
JPlagJava	2024/2025	79.61	46.62	13.05	1.17	0.19	0.00	0.00	0.00
JPlagJava	2025/2026	89.31	49.27	10.41	0.98	0.00	0.00	0.00	0.00
SIMGruneJava	2023/2024	17.95	2.94	0.60	0.23	0.08	0.08	0.08	0.08
SIMGruneJava	2024/2025	38.44	10.71	1.69	0.26	0.00	0.00	0.00	0.00
SIMGruneJava	2025/2026	30.05	5.52	0.70	0.00	0.00	0.00	0.00	0.00

Note: Values represent the percentage of all pairwise similarity scores at or above each threshold. Thresholds from 15% to 50% in five-percentage-point increments were examined to assess whether the observed patterns depend on the particular fixed thresholds used in the main analysis.

Table S5. Top-100 similarity score analysis

Tool	Assignment	Year	n	mean	median	minimum	maximum
JPlagJava	A1	2023/2024	100	32.3	31.4	28.4	51.9
SIMGruneJava	A1	2023/2024	100	20.9	19.5	15.5	55.0
JPlagJava	A1	2024/2025	100	30.3	29.4	26.8	44.5
SIMGruneJava	A1	2024/2025	100	23.9	22.5	20.0	35.0
JPlagJava	A1	2025/2026	100	28.8	28.0	26.1	38.1
SIMGruneJava	A1	2025/2026	100	20.2	19.5	18.0	33.5
JPlagJava	A2	2023/2024	100	31.0	30.2	28.4	60.1
SIMGruneJava	A2	2023/2024	100	20.4	19.0	17.0	58.5
JPlagJava	A2	2024/2025	100	28.6	28.0	26.7	38.8
SIMGruneJava	A2	2024/2025	100	23.7	23.0	21.0	33.0
JPlagJava	A2	2025/2026	100	27.9	27.3	25.9	32.6
SIMGruneJava	A2	2025/2026	100	21.7	21.0	19.5	29.5

Note: The analysis was performed on the original pairwise similarity scores. For each assignment-year-tool group, the 100 pairs with the highest similarity scores were selected. “Minimum” denotes the lowest similarity value among the selected top 100 pairwise scores.

Table S6. Cross-tool rank association between JPlagJava and SIMGruneJava.

Assignment	Academic Year	n submissions	n pairs	Spearman ρ	QAP p-value
A1	2023/2024	52	1326	0.600	<0.001
A1	2024/2025	59	1711	0.689	<0.001
A1	2025/2026	55	1485	0.572	<0.001
A2	2023/2024	52	1326	0.610	<0.001
A2	2024/2025	56	1540	0.682	<0.001
A2	2025/2026	54	1431	0.517	<0.001

Note: Spearman's rank correlation coefficient was calculated using matched pairwise similarity scores produced by JPlagJava and SIMGruneJava. Statistical significance was assessed using a two-sided QAP node-label permutation test with 9,999 permutations. QAP preserves the dyadic dependence structure resulting from individual submissions appearing in multiple pairwise comparisons.

Figure S8. Association between JPlagJava and SIMGruneJava pairwise similarity scores across assignments and academic years. Each panel compares similarity scores produced by the two tools for the same submission pairs. The upper row shows Assignment A1 and the lower row Assignment A2. Spearman's rank correlation coefficient is reported for each assignment-year group.

