Ten-Task Student Assessment Quality Project

Notes for SHS • Topic 83 • Fictional data only • 25 November 2026

Task 1 — Inspect raw records

Verify 252 imported rows and 240 unique business keys.

Raw input; profile fields and types; find duplicates by composite key.

Task 2 — Calculate missingness

18/252 = 7.1429% missing scores in imported raw rows.

Treat blank and N/A as missing only in AssessmentScore.

Task 3 — Detect duplicate observations

12 excess duplicates, 24 rows involved in repeated-key groups.

Business key StudentCode + AssessmentID; quarantine appended repeats.

Task 4 — Validate numeric fields

Eight out-of-range scores; investigate maximum-score rules.

Do not silently cap 105 or negative scores to bounds.

Task 5 — Review categories and dates

Ten nonstandard StudyLevel labels, seven invalid/ambiguous date entries.

Use source or an approved mapping, do not guess ambiguous dates.

Task 6 — Investigate statistical outliers

Median-of-halves IQR example flags 40 beyond upper fence 15.

A statistical flag is a review prompt, not deletion proof.

Task 7 — Write cleaning policy

Restore six score gaps from confirmed synthetic source, keep twelve unresolved.

Make evidence-level distinctions, preserve zeros and valid extremes.

Task 8 — Apply cleaning and log changes

Produce ready 240 rows, issue register 69 findings, 12 duplicates quarantined.

Every issue is classified as corrected, quarantined, retained valid or unresolved.

Task 9 — Recalculate metrics

Ready 240 rows, 12 missing, zero invalid scores/dates/categories/duplicate keys.

Compare percentages with stated 252 versus 240 denominators.

Task 10 — Report findings

Document 41 corrected, 12 quarantined, four retained valid, 12 unresolved.

Limit analyses requiring observed score to 228 ready rows; disclose limits.