

Solutions

2.1 Completely correct four-item sorting; liking/preference.

2.2 X: zero complete successes. Y: one complete success. Three partial correct placements are not three complete successes.

2.3 Award B1 each for three distinct controls: same wording, same item choices/key, same two-minute timing, no coaching, one layout per person. Assignment is a separate allocation procedure; do not repeat timing twice.

2.4 Allocating by reading confidence introduces a group difference tied to condition. Explain stopping/withholding before submission and inability to identify a sheet after anonymous submission. Pilot wording comprehension or workable timing after approval.

3.1 Award 2 per claim: correction, reason. Assignment within class volunteers is not school-wide random sampling. Measure the sorting task: liking is a different outcome. Keep the pilot separate and document changes: different instructions/procedural rehearsal are not the same main comparison.

3.2 Local analytic rubric, $8 \times$ B1: accuracy/preference distinction; same task/key; complete-success rule; repaired assignment; truthful anonymous-data limit; approval and separate pilot status; local-sample limit; coherent truthful 80–110 word response. Example: “Our question concerns completely correct sorting, so a liking survey cannot supply the primary result. We will use the same four-item task and fixed key for both layouts, with a two-minute limit and no coaching. A sheet counts as one success only when all four placements are correct. We will use shuffled assignment rather than place confident readers in B. Participants may stop or withhold a sheet; after anonymous submission we cannot reliably retrieve an individual sheet. Teacher approval is still pending. The later pilot will check wording and timing, separately from main data. Class volunteers will not represent every student.”