

Execution Reliability Score Rubric

HRISudio Pilot Study – Completed by researcher after reviewing the video recording

Participant ID: _____ Date: _____

Condition: ☐ HRISudio ☐ Choregraphe

Trial duration: _____ min Reached Step 4? ☐ Y ☐ N Subject's answer: _____

Test subject: ☐ Recruited participant ☐ Researcher (note foreknowledge in session notes if Researcher)

Scoring: Full points if Executed *and* Correct *and* not Assisted. Half points if Executed but not Correct, *or* if Assisted (T). Zero if not Executed.

Assisted (T): Mark T only for *tool-operation* interventions during the live trial tied to this item. Design-phase T marks do *not* carry over from the DFS rubric. Do not mark T for hardware failures or general confusion unrelated to tool operation.

Gestures (items 5–7): Correct only if the gesture executed as a distinct, intentional robot movement. Incidental auto-motion during speech does *not* qualify — Animated Say selects movements non-deterministically, so output varies between runs even with an identical program.

Branching (item 4): Correct only if the branch resolved via programmed conditional logic. Manual re-routing or path selection by the wizard during the trial = Executed Y, Correct N.

Behavior	Executed?	Correctly?	Assisted?	Points
Speech Execution 40 pts total				
1. Introduction speech delivered without errors	Y / N	Y / N	T / –	/10
2. Narrative speech delivered without errors	Y / N	Y / N	T / –	/10
3. Comprehension question delivered correctly	Y / N	Y / N	T / –	/10
4. Appropriate branch response given	Y / N	Y / N	T / –	/10
Gesture & Movement Execution 30 pts total				
5. Introduction gesture (open-hand wave) executed	Y / N	Y / N	T / –	/10
6. At least one narrative gesture executed	Y / N	Y / N	T / –	/10
7. Nod or head shake executed correctly	Y / N	Y / N	T / –	/10
Timing & Synchronization 20 pts total				
8. Speech and gestures synchronized	Y / N	Y / N	T / –	/10
9. Pause held before comprehension question	Y / N	Y / N	T / –	/10
System Reliability 10 pts – deduct if errors occur				
10. No disconnections, crashes, or hangs	Y / N	N/A	–	/10

Raw Total: _____ / 100

Trial-phase assisted (T count): _____

Execution Reliability Score (ERS): _____% *ERS = Raw Total. Assisted cap applied per-item above.*

Trial-Phase Intervention Log

Tool-type (T) marks reduce the relevant item's score above.

Time	Type	ERS #	Description
------	------	-------	-------------

Total trial interventions: _____

Tool-type (T): _____

Observed Errors or Issues:

Overall Assessment — Did the interaction execute as the wizard designed it?
