

Evaluate the evidence. Then evaluate the workflow.

For agencies, consulting firms, and research teams adding AI-moderated interviews to a real research brief.

Brief / client	Reviewer / date	Guide version	Sample / language

Mark each: P = passes this brief | R = revise and retest | F = does not fit. Circle essential checks first.

Evaluation criterion	Evidence to inspect	Status
01 Guide adherence	Does the conversation answer the learning goals, with the agreed guide preserved?	P / R / F
02 Useful, neutral probing	Do follow-ups clarify real experiences without supplying the answer or over-probing?	P / R / F
03 Modality and experience	Does the actual device and voice, video, or chat flow work for these participants?	P / R / F
04 Sample fit and quality	Are eligibility, recruitment source, exclusions, and quality checks documented?	P / R / F
05 Analysis and traceability	Can you trace a finding to complete evidence and retain contradictory cases?	P / R / F
06 End-to-end timing	What are setup, recruitment, review, and reporting time – not just interview time?	P / R / F
07 Complete project cost	Include platform, sample, incentives, commitments, and researcher delivery hours.	P / R / F
08 Language and meaning	Has a fluent reviewer checked the target language, terminology, and transcription?	P / R / F
09 Data and client boundaries	Do documented access, retention, deletion, and processing terms meet this brief?	P / R / F
10 Reuse across projects	Can you retrieve earlier evidence with its source, date, sample, and access limits?	P / R / F

Decision: Proceed / Revise / Use another method. A failed essential check cannot be averaged away.

Keep an evidence log: requirement, call or artifact, finding, unresolved issue, owner, and next step.

What does a useful follow-up actually do?

Inspect the published User Intuition phone-purchase call. Exchange numbers count displayed speaker turns, starting with the greeting.

Open the complete calls and readout: userintuition.ai/preview/

EXCHANGES 05-08

Turn a broad reason into a concrete event.

The moderator asks what triggered the purchase, then asks for the sequence. The participant describes a phone breaking while cooking.

Reviewer: Check whether the detail helps answer the brief.

EXCHANGES 09-12

Notice when a prompt may suggest a frame.

An either/or prompt contrasts a specific phone with getting online quickly, then explores the chosen device. Useful detail follows.

Reviewer: Test a more open prompt before accepting the frame.

EXCHANGES 13-20

Connect the decision to the experience.

Store choice leads to proximity, urgency, and staff interaction. The discussion goes beyond a broad satisfaction label.

Reviewer: Trace the interpretation back to the full exchange.

EXCHANGES 27-28

Leave room for criticism.

After positive comments, the moderator asks what could improve. The participant suggests digital signs.

Reviewer: Look for missed opportunities and negative cases too.

Keep the contrast. The phone buyer values staff help; the grocery shopper prefers shopping independently. Preserve the context instead of turning both calls into one universal recommendation.

Limits: This is a review of a visible example, not a controlled validation. The full original guide is not shown. Listen when wording matters: transcripts can contain errors. The 43-participant sample readout does not establish population prevalence.