Center for Assessment

Human-Guided Generative AI Across the Survey Lifecycle: Supplement

This is a companion to the CenterLine blog post "Human-Guided Generative AI Across the Survey Lifecycle" by André A. Rupp. Examples are illustrative and anonymized, drawn from two recurring project types: a district-level review of a balanced assessment system, and a state-level review of an accountability system.

Task	Assessment System Example	Accountability System Example	Human Guardrail
FRAMING & DESIGN			
Logic model / theory of action drafting	AI proposes an initial logic model linking district assessment practices (e.g., interim data use) to student learning goals.	AI proposes an initial theory of action linking state accountability indicators to district support needs.	Draft is a starting point only; Center staff sharpen it first, then the client iterates until it reflects their actual goals.
Research question drafting & refinement	AI drafts candidate questions (e.g., "How do teachers use assessment data to adjust instruction?") from early client conversations.	AI drafts candidate questions (e.g., "What do district leaders see as barriers to acting on ratings?").	Internal expert review sharpens framing first; client iteration confirms the questions match what they actually want to learn.
QUESTION DEVELOPMENT			
Response scale tradeoffs	AI compares a 4-point agreement scale vs. a frequency scale for an item on classroom assessment use.	AI compares a satisfaction scale vs. a trust scale for an item on accountability communication.	Item writer decides based on what the construct calls for, not AI's stylistic default.
Construct sharpening	AI helps distinguish "assessment literacy" from "assessment use" so an item targets the intended construct.	AI helps distinguish "trust in ratings" from "understanding of ratings" as separate constructs.	Human confirms the sharpened construct still matches the research question it's meant to serve.
Audience / age-appropriate language	AI simplifies a student-facing item about test feedback to a middle-school reading level.	AI adjusts a community-member item to avoid technical accountability jargon.	Reading-level target is set by someone who knows the respondent group; results are spot-checked.
Length / response burden	AI flags a teacher form running long relative to typical completion-time targets.	AI flags redundant items across sections of an accountability form.	Coverage-vs-burden tradeoff is a judgment call made by the project team, not the AI.
Bias, sensitivity, & inclusivity review	AI flags an item assuming home internet access when asking about take-home assessment prep.	AI flags language that may read differently to urban vs. rural community members.	Flag is a prompt for human review, not an automatic edit; the team decides how (or whether) to revise.
Distractor / response-option generation	AI proposes a fuller option set for "primary reason you don't use assessment data" before pretesting narrows it.	AI proposes a fuller option set for "main barrier to acting on accountability results."	Final option set is trimmed and validated using pilot feedback, not AI judgment alone.

Center for Assessment

Task	Assessment System Example	Accountability System Example	Human Guardrail
Codebook / rubric co-development	AI drafts a starter coding frame for open-ended comments on assessment quality.	AI drafts a starter coding frame for open-ended comments on accountability communication.	Frame is tested against real pilot responses and revised by a human coder before use at scale.
CONSTRUCT COVERAGE CROSSWALK			
Item ID tagging / database construction	Items tagged (e.g., T_S1_Q3 = Teacher, Section 1, Question 3) across student, teacher, and leader forms.	Items tagged similarly across district-leader, teacher, and community forms.	Tagging convention is set and checked by the team so IDs stay meaningful and consistent over time.
Cross-form coverage table assembly	AI builds a table showing which forms cover “use of data for instruction.”	AI builds a table showing which forms cover “trust in ratings.”	Human reviews the table for gaps the AI might miss (e.g., a construct assumed but never directly asked).
Wording consistency audit	AI flags a construct worded slightly differently on the teacher form vs. the leader form.	AI flags scale-label differences between two respondent-group forms.	Human decides whether the difference is intentional adaptation or accidental drift.
Client-facing sign-off package	Crosswalk, research questions, and form drafts bundled for district review.	Same bundle assembled for state client review.	Team confirms the bundle is accurate and complete before it goes to the client.
PILOT / COGNITIVE INTERVIEWS			
Session summarization	AI summarizes teacher think-aloud sessions into key takeaways on an assessment-use item.	AI summarizes community-member sessions into key takeaways on an accountability-ratings item.	Focus group leaders review summaries against original recordings or notes for accuracy before acting on them.
Revision recommendations	AI proposes a rewording based on repeated confusion about “formative use.”	AI proposes an added response option based on gaps stakeholders named.	Item writers evaluate the recommendation rather than auto-applying it.
Post-pilot crosswalk re-check	Any item revised after piloting triggers a fresh construct-coverage pass.	Same re-check applied after accountability-form revisions.	Design lead confirms coverage still holds.
DATA COLLECTION			
Outreach message drafting	AI drafts a sample email inviting teachers to complete the assessment-system survey.	AI drafts a sample message inviting community and business leaders to participate.	Client reviews and adapts messaging tone before it goes out under their name.
Outreach wave-scheduling artifact	AI drafts a table of who sends what to whom, by when, and at what frequency.	Same approach, scaled to a statewide respondent list.	Project team confirms the schedule is realistic and matches the actual survey window.

Center for Assessment

Task	Assessment System Example	Accountability System Example	Human Guardrail
Consent / sign-off language drafting	AI drafts consent language for the district-level survey.	AI drafts consent language for the state-level survey.	Client reviews language against relevant IRB or client requirements by a human.
ANALYSIS			
Cross-form variable merging	Seven versions of the same item across seven forms merged into a single analyzable variable.	Same technique applied to accountability-rating items across respondent-group forms.	Data analyst spot-checks merged variable against the source files before use in any analysis.
Thematic extraction (single item)	AI extracts themes from open-ended comments on assessment quality.	AI extracts themes from open-ended comments on accountability communication.	Themes are validated against a sample of raw responses by a data analyst or subject-matter expert.
Cross-question thematic synthesis	AI notices two different open-ended questions converging on the same concern about test-prep time.	AI notices convergence across questions on communication clarity.	Human confirms the “convergence” is a genuine pattern, not an artifact of similar wording.
REPORTING & DISSEMINATION			
Starter slide deck	AI drafts a deck summarizing district assessment findings for a school board.	AI drafts a deck summarizing statewide findings for state leadership.	Deck gets a human pass for context, framing, and narrative flow before use.
Brief / report drafting	AI drafts a full report on the district assessment review.	AI drafts a full report on the state accountability review.	Findings and interpretations are checked against underlying data by the project team.
Compact client summary	AI derives a short summary from the full district report.	AI derives a short summary from the full accountability report.	Summary is checked against the full report to confirm nothing material was lost or distorted.
Dashboard / visualization generation	AI drafts chart or dashboard specs summarizing assessment-use patterns.	AI drafts similar specs for accountability-rating patterns.	Visualizations are checked for accurate representation of the underlying data before sharing.
Accessibility formatting	AI helps format the district report with alt text and a plain-language summary.	Same applied to the state accountability report.	Final formatting is reviewed by the Center's in-house accessibility specialist.
Internal best-practice documentation	AI drafts internal guidance for replicating the district assessment review process.	AI drafts internal guidance for replicating the accountability review process.	Guidance is reviewed by senior staff before being adopted as a team practice.
Prompt / item database maintenance	Finalized items added to a searchable database with quality metadata (e.g., response-rate issues).	Same for accountability-review items.	Metadata and quality flags are reviewed by a human before items are reused in future projects.

Center for Assessment

Example Section of Construct Coverage Crosswalk

Construct Aspect	Coach	Interventionist	Test Coordinator	District Leader	School Leader	Teacher
Confidence explaining assessment purposes/uses to relevant audiences	C2_Q2	I2_Q1	TC1_Q1	D1_Q1	S1_Q1	T1_Q1
Perceived accuracy / construct validity of assessments	C2_Q1	I2_Q2	TC1_Q2	D1_Q2	S1_Q2	T1_Q3
Confidence completing assessment-related tasks (e.g., locate data, interpret results, use data)	C2_Q3	I2_Q3	—	—	—	T1_Q2

Note. Structure of ID mirrors design. For example, C2_Q2 is the second question in the second section on the survey form for coaches.