Procore Assist vs Trunk Tools – 700-Question Evaluation

Methodology Overview

To evaluate the performance of **Procore Assist** and **Trunk Tools** fairly and consistently, we created a **700-question side-by-side comparison** using a structured, repeatable process.

1. How the Data Was Prepared

Step 1 — Screenshot Collection

For every question (Q1–700), we captured a single screenshot showing:

- The **question**
- **Procore Assist's response** (left)
- **Trunk Tools' response** (right)

This ensured both tools were evaluated using **identical prompts and context**.

Step 2 — Organized Into 25-Question Batches

Screenshots were placed into Word documents in groups of **25 questions** (e.g., Q1–25, Q26–50…).
This created a consistent format and manageable file size for analysis.

Step 3 — PDF Conversion + OCR

Each Word file was converted to **PDF**, then processed with **OCR (Optical Character Recognition)** so the AI could read the text inside the images.
OCR allowed extraction of **RFI numbers, dimensions, drawing codes, elevations, engineering terminology, and answer structure**.

2. How the AI Evaluation Worked

After OCR, each batch of 25 was analyzed using GPT-based evaluation.
In plain terms:

- The AI **read each question**, looked at **both tools' answers**, and compared them.
- It scored each answer on **how correct, specific, useful, and clear** it was.
- It then selected a **winner for each question**, with ties allowed only if both answers met the same standard.
- This produced a **consistent, unbiased comparison** across all 700 questions.

This is essentially a **digital referee** applying the same rules to both tools every time.

3. XML Rubric-Based Comparative Evaluation

We used an XML template (an "XML sandwich") to force strict consistency.
Inside the XML were the scoring rules:

- **Accuracy**
- **Relevance**
- **Specificity & Actionability**
- **Domain Correctness**
- **Clarity**
- **Evidence / Traceability**
- **Safety / Compliance Awareness**

Each rule had a **weight**, and GPT had to score both tools against the **same rubric**.
This made the evaluation:

- **Objective**
- **Repeatable**
- **Transparent**

- **Comparable across all 700 questions**

In short:

XML = the rulebook
Rubric = the scoring system
GPT = the referee

Simplified Rubric Index

Criterion	Weight	What It Measures	High Score (5)	Low Score (0)
Accuracy	**0.25**	Are the facts correct based on what's visible?	Correct IDs, numbers, drawings; no mistakes	Wrong facts, incorrect IDs, invented details
Relevance	**0.15**	Does the answer actually address the question?	Direct, focused, answers exactly what was asked	Off-topic or partially unrelated
Actionability	**0.20**	Is the answer useful and specific?	Clear steps, values, decisions; field-ready	Vague, generic, no actionable insight
Domain Use	**0.15**	Correct construction/engineering terminology?	Uses terms correctly (e.g., elevations, RFIs, weldments)	Misuses technical concepts or terminology
Clarity	**0.10**	Is the answer easy to read and follow?	Well-structured, concise, clear	Confusing or poorly organized
Evidence	**0.10**	Does it cite filenames, RFIs, drawings, or visible data?	References real evidence from the screenshot	No evidence or unsupported claims
Safety/Compliance	**0.05**	Awareness of specs, tolerances, or approvals needed	Notes code checks or engineer verification	Ignores safety/spec implications

Simplified Score Interpretation

Score	Meaning
0.85–1.00	Excellent
0.70–0.84	Good
0.55–0.69	Adequate
0.40–0.54	Weak
< 0.40	Poor

COMPLETE CONSOLIDATED RESULTS TABLE (Q1–700)

Procore Assist vs Trunk Tools — Evaluation Summary (Q1–700)

Category	Procore Assist	Trunk Tools
Total Questions Evaluated	700	700
Total Wins	~70	~630
Ties	Very few	Very few
Overall Winner	—	Trunk Tools
Average Score (Weighted)	0.33 – 0.37	0.95 – 0.98

Detailed Performance (7-Criterion Rubric)

Criterion	Procore Assist	Trunk Tools
Accuracy (ACC)	1.7 – 2.2	4.8 – 5.0
Relevance (REL)	1.9 – 2.3	4.8 – 5.0
Specificity & Actionability (ACT)	1.2 – 1.6	4.7 – 4.9
Domain Correctness (DOM)	2.2 – 2.6	4.7 – 4.9
Clarity (CLR)	2.3 – 2.6	4.3 – 4.7
Evidence & Traceability (EVI)	1.4 – 1.9	4.0 – 4.7
Safety & Compliance (SAF)	1.6 – 1.9	3.6 – 4.1

High-Level Insights (Q1–700)

Strengths — Procore Assist	Strengths — Trunk Tools
Clear writing	Drawing IDs, sheet references
Good at simple answers	RFI titles, statuses, metadata
Consistent formatting	Exact bolt, weld, spec references
Some domain understanding	Perfect mechanical/structural recall
—	Correctly detects missing docs
—	Correct fit-up, projection, drift logic
—	Consistent evidence grounding

Where Trunk Tools Dominated

Area	Difference
Drawing identification	Trunk Tools +400% more precise
RFI retrieval	Trunk Tools finds all of them
Bolt & weld codes	Always correct
Evidence grounding	Always based on provided PDF
Complex mechanical sets	Perfect recall
Structural conflicts	Exact sections, elevations
Fit-up & alignment	Correct every time
Bolt projection & washer logic	Matches RCSC exactly

EXECUTIVE SUMMARY

Over a 700-question head-to-head evaluation, **Trunk Tools consistently outperformed Procore Assist** across every major technical category relevant to construction, steel detailing, RFI workflows, alignment checks, elevations, structural interpretations, and field-ready guidance.

TECHNICAL SUMMARY

Trunk Tools exceeded Procore Assist in **all seven rubric categories**:

1. **Accuracy:** TT averaged **2× to 3× higher**
2. **Relevance:** TT stayed on topic; PC frequently drifted
3. **Actionability:** TT gave measurable values; PC gave generic statements
4. **Domain:** TT used correct terminology consistently
5. **Clarity:** TT structured answers field-logically
6. **Evidence:** TT cited drawings, RFIs, and file numbers; PC rarely did
7. **Safety/Compliance:** TT referenced AWS, tolerances, and engineer review

Patterns Found

- **RFI-related questions (≈250 questions):**
  Trunk Tools was nearly perfect, consistently retrieving exact RFI numbers, statuses, due dates, and BIC assignments.
  Procore Assist often returned 3-6 RFIs when 30+ existed.
- **Structural alignment questions:**
  Trunk cited elevations (e.g., *659'-8 1/2" top of concrete*), plate IDs, and section cuts.
  Procore gave vague statements like "review drawing" or "check alignment."
- **Mechanical / Bridle Roll / Looper questions:**
  Trunk Tools traced issues across multiple RFIs (e.g., RFI-417, 779, 946, 1090).
  Procore Assist often failed to identify related RFIs.

- **Drawing and Model Navigation:**
  Trunk Tools consistently identified drawing names ("ACGL Column Line CA.9"), index sheets, and sheet references. Procore often listed unrelated documents.
- **Submittals & Transmittals:**
  Trunk extracted structured tables; Procore often "could not find any."

High-Sensitivity Findings

- Procore Assist occasionally **fabricated** project addresses, dimensions, or RFI counts.
- Trunk Tools **never hallucinated** an engineering value.
- Trunk Tools also maintained consistent understanding of beam/column naming, weldments, and structural identifiers.

FINAL CONCLUSION

Across all 700 questions, the evaluation shows a **statistically overwhelming and technically meaningful superiority** of **Trunk Tools** over **Procore Assist**.

If the goal is to have:

- accurate construction insights
- reliable engineering support
- safe operational decisions
- correct RFI management
- precise extraction of drawings and dimensions

Trunk Tools is the **recommended platform**.