nine twothree

METHOD · VERSION 1.0



The Shop Context Test

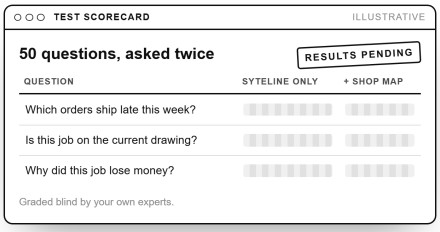
A repeatable way to test whether the same AI model answers better when it knows your shop. Your people write the questions, and your experts score the answers.

PURPOSE

The claim we are testing

We think frontier AI models already know manufacturing. What they lack is knowledge of one shop: its rules, exceptions, unofficial sources and the reasons behind its numbers. The Shop Map captures that knowledge.

If we're wrong, the test will show it, and we'll publish that too.



Example scorecard layout. No test has finished, so there are no results.

Five steps

1 WRITE AND FREEZE THE QUESTIONS
The shop writes 50 questions before Shop Map work starts. Frozen questions are never edited.

3 RUN A
Run the model twice. Same model, same settings, same SyteLine access and tools. No Shop Map.

3 RUN B
Identical to Run A, plus the Shop Map.

Run B only

2 BUILD THE SHOP MAP INDEPENDENTLY
Harness sessions run as normal. The map builders never see the frozen questions.

4 SCORE BLIND
Answers are shuffled and labelled X and Y. Two of the shop's experts score each answer.

5 RECORD EVERY MISS
Every wrong or unsourced answer goes into a miss log with its category and the reason it failed.

PROTOCOL

How the questions are written

Each question is written the way it's asked out loud, with the expected answer and the source the expert would use.

CATEGORY	QUESTIONS	EXAMPLE OF THE KIND OF QUESTION
Scheduling and delivery	10	Which jobs will miss their ship date this week, and what should move?
Job cost and margin	8	Why did job margin on this part family drop last quarter?
Quality and revisions	8	Which open jobs were released before the latest engineering change?
Purchasing and suppliers	8	Which late purchase orders put a customer order at risk?
Inventory and materials	8	Which substitute material has this customer approved before?
Quoting and history	8	What did we charge last time for a part like this, and did we make money?

How the runs are kept fair

- Nobody tunes prompts against the frozen questions
- Run order and settings are logged and kept
- The shop's experts score; NineTwoThree does not

How answers are scored

SCORE	SCALE	WHAT THE EXPERT CHECKS
Correct	0, 1 or 2	2 is right and complete. 1 is partly right or missing a step that matters. 0 is wrong or no answer.
Sourced	0 or 1	1 if the answer names the records or documents it used and they support the answer.
Time	Minutes	Time from question to a usable answer, compared with how long the expert says it takes today.

Where the two experts disagree on Correct, a third decides. The miss log records the run and why each answer failed: wrong source, missing rule, stale data, misread question or a model error.

WHAT THE SHOP GETS

- The full scorecard for both runs, by question and by category
- The Shop Map entries built during the test
- A plain summary for the owner and the leadership team

PUBLICATION RULES

- Results are published only after a test is complete
- Nothing is published without the shop's written approval
- The shop can stay anonymous
- Every miss is published along with the wins
- No projected or estimated results, ever

Results

Results in progress This table will be filled only with completed, approved results.

CATEGORY	RUN A CORRECT	RUN B CORRECT	RUN A SOURCED	RUN B SOURCED	MEDIAN TIME A	MEDIAN TIME B
Scheduling and delivery	In progress	In progress	In progress	In progress	In progress	In progress
Job cost and margin	In progress	In progress	In progress	In progress	In progress	In progress
Quality and revisions	In progress	In progress	In progress	In progress	In progress	In progress
Purchasing and suppliers	In progress	In progress	In progress	In progress	In progress	In progress
Inventory and materials	In progress	In progress	In progress	In progress	In progress	In progress
Quoting and history	In progress	In progress	In progress	In progress	In progress	In progress

Taking part

It suits a SyteLine shop where at least six people answer operational questions every week and an ERP owner can grant read access. It takes the shop about 10 hours to write and score questions, over six to eight weeks, alongside a Shop Harness engagement.

Put your shop in the test

Book a Shop Walk at ninetwothree.co and tick "Shop Context Test".

Book a Shop Walk

Infor, CloudSuite Industrial and SyteLine are trademarks of Infor. NineTwoThree is an independent company and is not affiliated with or endorsed by Infor. Copyright © 2026 923 Digital, LLC.