
INDEPENDENT EVALUATION

IGCE Builder FFP Testing Record

Firm-Fixed-Price cost estimate skill validated across fifteen runs in three testing waves on two Claude models; Rounds 5 and 6 patches shipped

Skill

igce-builder-ffp (post-Round 6 patches; 42/42 Wave 3 retest assertions passed on Opus)
github.com/1102tools/federal-contracting-skills

Date

April 2026
1102tools.com

Part 1: For Federal Acquisition Users

The bottom line

Three waves of independent testing across April 2026 (15 end-to-end runs on Claude Opus 4.7 and Claude Sonnet 4.6, 210 binary assertions graded) show the IGCE Builder FFP skill reliably produces auditable Firm-Fixed-Price cost estimates across four distinct federal acquisition scenarios. Wave 1 surfaced 17 cross-run quality issues, all patched. Wave 2 validated the patches. Wave 3 re-tested post-Round 4 substrate patches, surfaced three burden-band gaps, shipped Round 5 + Round 6 patches, and closed clean at 42/42.

Wave 1: 55/56 = 98% (Opus) with Sonnet parity. Wave 2: 56/56 = 100% (Opus) after 17 substrate patches. Wave 3: 42/42 = 100% (Opus) after Round 5 + Round 6 patches.

Scenarios tested and how reliably they work

Scenario	Models	Result
Standard FFP-by-period IT services build (DC dev team, base + 2 OYs)	Opus, Sonnet	Reliable both waves
24x7 shift coverage (Cleveland SOC analysts, base + 2 OYs)	Opus, Sonnet	Wave 1: Opus burned 15 min brute-forcing Cleveland MSA due to 2024 OMB renumbering from 17460 to 17410, shift math reconstructed from first principles. Wave 2: reliable after patches
Physical engineering multi-LCAT (Oak Ridge DOE environment)	Opus, Sonnet	Wave 1: SOC mapping defaulted to IT codes for Mechanical/Electrical Engineers, PM mapped to 11-3021. Wave 2: reliable after 17-2xxx block added and PM SOC made context-dependent
FFP-by-deliverable multi-milestone study (DISA feasibility analysis)	Opus, Sonnet	Wave 1: hour allocation across deliverables inconsistent (4/4 runs chose different methods), CALC+ rate validation hit silent-wrong-answer bug (q= vs keyword=) in 4/4 Opus runs. Wave 2: reliable after explicit allocation methods + CALC+ query signature inlined

Manual-verification checklist

Scan every output for these before using in a contract file:

1. CALC+ rate validation was actually validated, not faked. The Wave 1 silent-wrong-answer bug came from the CALC+ API accepting q= and returning the full 265K-record corpus without error. If the rate validation sheet cites a median derived from "hundreds of matches" or variance exceeds 60% across the sample, the worker hit the bug. Patched skill inlines keyword= and /v3/api/ceilingrates/ endpoint; workers cannot accidentally route through the broken signature.

2. Fully burdened rate must tie to BLS + wrap buildup, not just to CALC+ directly. CALC+ reports awarded ceiling rates for completed MAS tasks. Using a CALC+ median as your direct FFP rate skips the buildup audit trail FAR 15.404-1(a) expects. Correct flow: BLS base → fringe

→ labor+fringe → overhead → subtotal → G&A → total cost → profit = FBR. Then compare FBR to CALC+ for reasonableness.

3. 24x7 coverage needs 4.2 FTE per seat, not 1. The skill's Step 0.5 computes 2,080 productive hours × 4.2 FTE to cover 8,760 annual hours (24 × 365) with leave/training/coverage slack. If a Cleveland SOC workbook shows 3 FTE for 24x7 coverage, the shift math is wrong. Double-seat (two analysts always on) is 8.4 FTE.

4. Rate validation band should be 0-40% above CALC+ median for FFP. Wave 2 patch calibrated: 0-15% is expected range. 15-40% is the FFP premium band (risk-adjusted fixed pricing justifies a markup over the ceiling-rate median). Above 40% needs explicit justification in the narrative. The pre-patch skill flagged anything over 10% which would have fired on nearly every legitimate FFP build.

5. MSA renumbering silently returns empty data. BLS returns the same "series does not exist" for a truly unpublished occupation and for a renumbered metro. If a metro query returns NO_DATA across every SOC, check the OMB Bulletin 23-01 renumbering list before falling back to state. Cleveland (17460→17410), and possibly Dayton, shifted.

6. Implied multiplier tells you whether the build makes sense. Fully burdened rate divided by BLS base wage should land in the 2.2x-3.5x band for MID scenario. Below 2.2x means an unrealistic wrap assumption. Above 3.5x needs SCIF/OCONUS/niche justification. The HIGH scenario legitimately exceeds 3.5x and should not be flagged.

Choosing between Opus 4.7 and Sonnet 4.6

Short answer: both work. Use either on the patched skill. Differences are small.

Opus 4.7 handles multi-location builds and FFP-by-deliverable decomposition more reliably. In Wave 1, Opus caught the Cleveland MSA renumbering by brute-force scanning BLS; Sonnet fell back to state-level wages without flagging the metro issue. Opus is the preferred model for: SOW-driven builds where decomposition judgement matters, FFP-by-deliverable with per-LCAT hour matrices, any build touching a DOE or DoD specialty metro, any multi-LCAT build with 5+ labor categories. Opus also tool-uses more aggressively and sometimes hits the per-response tool-use cap on complex builds. If that happens, click continue.

Sonnet 4.6 is faster on standard single-location FFP-by-period builds and produces cleaner workbooks in fewer tokens on the happy path. For a straightforward "3 FTE in DC, base + 2 OYs" job, Sonnet wins. Sonnet is less reliable at: metro-code validation when BLS returns empty, FFP-by-de-

liverable hour allocation decisions, and rate validation narrative text (tends to state rather than justify).

Wave 2 was run on Opus only for time reasons; Sonnet parity on the patched skill is inferred from Wave 1 where Sonnet matched Opus on 3 of 4 scenarios.

What the skill does not do

- **It does not produce LH/T&M or cost-reimbursement estimates.** Use IGCE Builder LH/T&M or IGCE Builder CR respectively. The wrap rate buildup in FFP does not apply to those contract types.
- **It does not estimate subcontractor costs.** If the prime proposes 30% subcontract, you need separate vendor input or a second IGCE for the sub's scope.
- **It does not negotiate fee/profit.** It produces a cost buildup. Fee negotiation is a separate FAR 15.404-4 activity.
- **It does not handle OCONUS per diem.** GSA Per Diem covers CONUS only; use State Department rates for OCONUS assignments.
- **It does not price SCIF build-out, TEMPEST, or COMSEC equipment.** These require agency-specific quotes.
- **It has not been tested on:** CR-to-FFP conversion modeling (pricing legacy cost-plus work as FFP), FFP with award fee overlays (hybrid structures), ANSI/EIA-748 EVMS-compliant cost buildup formatting, international labor (BLS is US-only), or DCAA forward-pricing rate audits (the skill estimates cost, it does not audit vendor rate proposals against DCAA-disclosed rates).

Environmental gotchas on claude.ai web chat

Gotcha	What happens	Workaround
Multi-LCAT build with 5+ categories + per-deliverable matrices	Opus hits per-response tool-use cap mid-build	Click continue; the skill resumes without repeating prior work
Complex workbook (9 sheets + rate validation dual-pool)	Python execution on claude.ai can time out	Ask model to build each sheet incrementally and present sheet-by-sheet
xlsx output doesn't appear in chat	File is in sandbox but not surfaced to UI	Patched skill's Step 9 explicitly calls <code>present_files()</code> after copying to <code>/mnt/user-data/outputs/</code>
CALC+ dual-pool query returns inconsistent record counts	Title-match and experience-match pools sometimes overlap	Use the patched Step 4 decision tree: title-match first, then experience-match as sanity layer if title match $N < 10$

Part 2: For Developers and Technical Reviewers

Testing methodology

Scenarios

Four scenarios were selected before any testing began, chosen to exercise distinct capabilities across FFP pricing structures and federal agency contexts:

- **S1 — Standard FFP-by-period IT services:** 3 FTE developer team in DC, base year + 2 option years. Exercises SOC mapping for Software Developer (15-1252) + Senior Developer + Business Analyst, BLS DC metro (47900), full wrap buildup (35% / 85% / 10% / 10% MID), 2.5% escalation year-over-year, CALC+ rate validation at 520 SIN.

- **S2 — 24x7 shift coverage, specialty market:** Cleveland SOC analyst coverage 24x7x365, base + 2 OYs. Exercises shift coverage math (single-seat = 4.2 FTE), Information Security Analyst SOC (15-1212) BLS lookup, Cleveland MSA boundary (exposes 2024 OMB renumbering from 17460 → 17410), rate validation for an above-median SOC in a below-median metro.
- **S3 — Physical engineering multi-LCAT:** Oak Ridge TN DOE environment, 6-category staffing (Mechanical Engineer, Electrical Engineer, Chemical Engineer, Technical Writer, PM, Admin), base + 4 OYs. Exercises 17-2xxx SOC block (Wave 1 missed: skill defaulted Mechanical Engineer to 15-1211), PM SOC context-dependent (should be 11-9041 Engineering Manager, Wave 1 picked 11-3021 IT Manager), Oak Ridge MSA (28940), multi-LCAT wrap buildup, CALC+ dual-pool rate validation on senior engineering LCATs.
- **S4 — FFP-by-deliverable, multi-milestone study:** DISA 18-month feasibility study, 4 deliverables at 15/30/25/30 scope weights, SOW-driven build (Workflow A+). Exercises Step 0 requirements decomposition + validation gate, FFP-by-deliverable hour allocation across milestones (three valid methods: uniform split, per-LCAT matrix, staffing-profile), aging wages once to contract start with no mid-contract escalation, Summary sheet columns = CLINs.

Each scenario had a 14-point binary assertion matrix. Assertions were written before any worker output was seen and were not revised after the fact.

Environment

- claude.ai web chat, fresh conversation per run
- Skills installed: `igce-builder-ffp` plus downstream `substrate-bls-oews-api`, `gsa-calc-ceilingrates`, `gsa-perdiem-rates` (all merged, post-Round 2 patches for downstream skills)
- Models: Wave 1 ran each scenario on Opus 4.7 and Sonnet 4.6. Wave 2 ran Opus 4.7 only (time-constrained)
- Total: Wave 1 = $4 \times 2 = 8$ runs. Wave 2 = $4 \times 1 = 4$ runs. Aggregate = 12 runs / 168 assertions

Grading

The grader (Claude Code session separate from any worker run) read only the worker's final response text and produced workbook. Workers were not coached during runs. Each assertion graded binary pass/fail. Suspicious details were noted even when assertions passed. Hour allocation ambiguity in S4 was graded as "worker picked one valid method and stayed internally consistent" rather than prescribing a specific method.

Wave 1 results (pre-patch)

Scenario	Sonnet 4.6	Opus 4.7
S1 DC dev team FFP-by-period	14/14	14/14
S2 Cleveland 24x7 SOC	13/14	14/14
S3 Oak Ridge DOE multi-LCAT	14/14	14/14
S4 DISA FFP-by-deliverable	14/14	13/14
Total	55/56 (98%)	55/56 (98%)

Wave 1 aggregate: 110/112 (98%).

Failures observed

S2.X Sonnet — Cleveland 24x7 shift math: Sonnet computed 3 FTE for single-seat 24x7 coverage. Correct is 4.2 FTE (8,760 annual coverage hours / 2,080 productive hours × availability factor for leave/training/turnover). The final workbook understaffed by 28% and the FFP total was commensurately low.

S4.X Opus — CALC+ rate validation returned meaningless results: Opus sent CALC+ queries using q= parameter. CALC+ accepted silently and returned the full 265K-record corpus. Rate validation narrative cited a "median of \$142.85 across 15,000+ matches" which was meaningless (population median, not occupation-specific). Workbook shipped with a broken validation sheet that looked fine on casual inspection.

The CALC+ bug was NOT unique to S4. Targeted re-inspection showed all 4 Opus Wave 1 runs hit this bug to varying degrees. S1, S2, S3 Opus runs still passed their rate-validation assertions because the grader checked "rate validation sheet exists with a median and a variance band" rather than "the median is arithmetically defensible." The patched assertion text for Wave 2 required the worker to cite the exact endpoint (/v3/api/ceilingrates/) and the exact parameter (keyword=) in methodology notes, forcing the query signature to be demonstrated rather than merely claimed.

Wave 1 findings: 17 cross-run issues patched

From Wave 1 worker self-assessments, grader notes, and cross-run observation:

1. **CALC+ query signature silently wrong in 4/4 Opus runs.** q= returns the full corpus; keyword= returns the filtered set. The CALC+ skill documentation showed q= in one example. The downstream skill was patched separately (see GSA CALC+ testing record). The FFP skill's Step 4 now inlines the correct endpoint, parameter name, and JSON path explicitly: no substrate lookup required.
2. **24x7 shift coverage math missing.** Workers reconstructed from first principles with varying results. Added Step 0.5 "Shift Coverage Staffing" with 4.2 FTE single-seat and 8.4 FTE double-seat formulas and worked example.
3. **Physical engineering SOCs absent from the mapping table.** Workers defaulted Mechanical Engineer, Electrical Engineer, etc. to 15-1211 (Computer Systems Analyst) or 17-2199 (Engineers, All Other). Added explicit 17-2xxx block: 17-2011 Aerospace, 17-2031 Biomedical, 17-2041 Chemical, 17-2051 Civil, 17-2071 Electrical, 17-2072 Electronics, 17-2081 Environmental, 17-2112 Industrial, 17-2141 Mechanical, 17-2161 Nuclear, 17-2171 Petroleum.
4. **PM SOC mapping conflated.** Workers defaulted Program Manager to 11-3021 (Computer and Information Systems Managers) regardless of context. Patched to context-dependent: 11-1021 General and Operations Manager (default / ops), 11-9041 Architectural and Engineering Manager (physical engineering programs), 11-3021 Computer and Information Systems Managers (IT programs only).
5. **Cleveland MSA renumbering not flagged.** Opus S2 burned significant time brute-force scanning. Patched in the downstream BLS skill (Round 3) and cross-referenced in FFP Step 2 with a silent-wrong-answer trap entry.
6. **BLS series ID component lengths not documented.** Workers constructed invalid 24- or 26-char IDs and retried. Added component breakdown: prefix(4) + area(7) + industry(6) + SOC(6) + datatype(2) = 25 chars total. Documented in Step 2 with a worked example.
7. **Seniority modeling absent.** Default wage pull was mean or median only. For Senior/Junior LCATs, workers needed interquartile context. Added P25 → Junior, P50 → Mid, P75 → Senior pattern in Step 2 with explicit instruction to pull all 5 percentiles.
8. **Aging factor hardcoded rather than cell-referenced.** Wave 1 workers applied aging as "× 1.023" hardcoded in formulas. If user changes the contract start assumption, the whole sheet recomputes wrong. Patched Step 2B + Step 8 to require cell-referenced formula: =BLS_2024_wage * ((1 + escalation)^months_gap_12) with BLS_vintage, contract_start, months_gap, and aging_factor as named assumption block rows (9-12).
9. **Rate validation flag band miscalibrated.** Pre-patch threshold was 10%; legitimate FFP premiums routinely exceeded that. Patched: 0-15% expected, 15-40% FFP premium band, >40% needs justification.

10. **CALC+ dual-pool analysis undocumented.** For senior LCATs, title-match alone often returns $N < 10$. Added dual-pool method in Step 4: title-match primary, experience-match secondary, report both counts and both medians.
11. **0-night day trip edge case missing.** Day trips (same-day return) use partial M&IE only, no lodging. Pre-patch Step 5 didn't distinguish. Added explicit 0-night case: 75% M&IE first day, no lodging, no last-day M&IE.
12. **"No travel" Sheet 5 handling absent.** Workbook always built Sheet 5 with zeros and placeholder text that broke downstream formulas. Patched: if travel = 0, Sheet 5 says "Travel Not Applicable" with no SUM references.
13. **Multi-location with explicit headcount triggered an unneeded prompt.** Workers asked "Option A (blend), B (lead location), or C (separate lines)" even when user gave per-location headcount. Patched: Option C default when headcount per location is explicit. Prompt only if blend is ambiguous.
14. **FFP-by-deliverable hour allocation method was user-choice with no guidance.** Workers picked differently across 4 S4 runs (uniform split, per-LCAT matrix, staffing-profile weighted). All three are valid; the skill didn't say so. Patched Step 7 with three methods documented, selection guidance by project size, and requirement for worker to cite which method they chose.
15. **Deliverable-timing escalation inconsistent.** Workers sometimes applied escalation within a single PoP, sometimes not. Patched: single-period PoP gets aging-to-start only, no mid-contract escalation. Multi-year PoP applies escalation to out-years per Step 7.
16. **Sheet 2 block layout formulas absent.** Workers built row references by hand for each LCAT block. Patched Step 8 with explicit formula: $\text{row}(N) = 1 + (N-1) * 19$, FBR at offset +17, multiplier at +18. Verifiable in a glance.
17. **No explicit final-step "present the file."** Workers wrote to `/mnt/user-data/outputs/` but sometimes didn't call `present_files()`. File existed in sandbox but wasn't surfaced to UI. Added Step 9: explicit copy-and-present pattern.

Bonus patches shipped alongside: - Annotation text cannot start with = + - @ (Excel formula parse). Documented in Step 8 with escape guidance. - ODC placeholders must be numeric 0 (not text "TBD") to prevent #VALUE! propagating through SUM formulas. Documented in Step 5. - Implied multiplier column handling when user doesn't want the audit column: drop or annotate as non-billable. - Domain triage first: the skill now instructs the worker to identify agency domain (DoD / IC / DOE / civilian IT / research) before SOC mapping. Domain signals which SOC block applies.

Wave 2 results (post-patch)

Scenario	Opus 4.7
S1 DC dev team FFP-by-period	14/14
S2 Cleveland 24x7 SOC	14/14
S3 Oak Ridge DOE multi-LCAT	14/14
S4 DISA FFP-by-deliverable	14/14
Total	56/56 (100%)

Wave 2 aggregate: 56/56 (100%). All 17 Wave 1 issues fixed; no new failures observed.

Methodology upgrades observed beyond the matrix

Wave 2 Opus workers produced stronger output even on assertions that passed in Wave 1:

- **S1:** used P25/P50/P75 for Junior/Mid/Senior variants explicitly; cited the patched FFP premium band (15-40%) in rate validation narrative; dropped implied-multiplier column with justification note.
- **S2:** used Cleveland 0017410 directly (no brute-force scan); computed 4.2 FTE via the Step 0.5 worked example; noted the 2024 OMB renumbering explicitly in methodology.
- **S3:** used 17-2141 Mechanical, 17-2071 Electrical, 17-2041 Chemical from the new engineering block; selected 11-9041 Engineering Manager as PM SOC with context justification; applied dual-pool CALC+ for senior engineers with title-match N=4 + experience-match N=27 both reported.
- **S4:** chose staffing-profile allocation with explicit rationale (matrix too complex for 6 LCATs × 4 deliverables, uniform split violated known back-loading of D3+D4); applied aging-to-start only (no mid-contract escalation on 18-month single PoP); cited /v3/api/ceilingrates/ + keyword= explicitly in CALC+ validation methodology; called `present_files()` in Step 9.

What was not tested

- Sonnet 4.6 on the post-patch skill (inferred from Wave 1 parity on 3/4 scenarios; not directly validated)
- FFP with award fee overlay (hybrid FFP + award fee structures)

- CR-to-FFP conversion modeling (pricing legacy cost-plus scopes as FFP)
- OCONUS travel CLINs (State Department rates)
- ANSI/EIA-748 EVMS-compliant cost buildup formatting
- DCAA forward-pricing rate proposal audits (distinct activity from IGCE build)
- Indefinite Delivery vehicles with seed FFP task orders (ordering-vehicle-level pricing)
- Contract bundling or consolidation scenarios with cross-location overhead pools
- International labor / EU wage data (BLS is US-only)
- Uncertainty quantification beyond the three-scenario band (Monte Carlo, sensitivity analysis)

Wave 3 retest (post-Round 4 substrate validation)

Three Opus scenarios re-run against the patched skill in April 2026, same 42-assertion matrix as Wave 2. Scenarios re-exercised Dayton MSA renumbering (now BLS-patched), DoE M&O overhead environment, GSA MAS commercial burden preset, o-night day trip, and the CALC+ dual-pool pattern.

Scenario	Wave 2	Wave 3 retest
S1 Wright-Patterson DoD Secret engineering	13/14	14/14
S2 Oak Ridge DOE FFP-by-deliverable	13/14	14/14
S3 NASA Glenn GSA MAS 24x7 SOC	13/14	14/14
Total	39/42 (93%)	42/42 (100%)

All three previously-failed burden-band assertions flipped to PASS after Round 5 patches shipped. Zero regressions on the 39 previously-passing assertions.

Round 5 patches shipped (between Wave 2 and Wave 3 retest)

1. **Wrap rate presets by contract vehicle** (10-row table) added to Information to Collect. Explicit instruction to ASK about contract vehicle before defaulting to skill mid. Covers GSA MAS commercial/cleared, Agency BPA non-cleared/cleared, DoD prime non-cleared/Secret/SCIF, DoE M&O/FFRDC, R&D CR, OCONUS.

2. **DATEDIF formula fix.** Replaced malformed $\text{YEAR}(\text{LEFT}(B9, 4))$ (Excel can't apply YEAR to a string) with $(\text{VALUE}(\text{LEFT}(B10, 4)) - \text{VALUE}(\text{LEFT}(B9, 4))) * 12 + \dots$. Every Wave 2 worker had to patch this in-place.
3. **"Wait - 19 rows" drafting artifact removed.** Sheet 2 block layout now reads clean.
4. **Cap decision tree extended for P75-also-capped case.** Knoxville Nuclear Sr / LANL physicist pattern. Skill now prescribes: use Mean when P75 caps, cross-reference commercial surveys, apply national P75/median ratio if deriving.
5. **Contract start date default.** Auto-default to October 1 of next federal fiscal year, surfaced as blue-font editable cell. No more silent invention.

Round 6 patches shipped (Wave 3 retest findings)

Independent workers in Wave 3 each caught the same math error in the Round 5 preset table's "Implied multiplier" column. Three workers independently computed the compounded arithmetic and flagged it.

1. **Preset multiplier column math corrected** across all 10 vehicle rows. Example: GSA MAS commercial 30/60/10/8 was "~1.9x" (wrong); actual math $1.30 \times 1.60 \times 1.10 \times 1.08 = 2.47x$. Corrected row-by-row. Added an explicit "Math check" line showing the compounding formula so builders can verify.
2. **Vehicle-aware sanity band.** The generic 2.2x-3.5x commercial band misfires against cleared DoD and DoE M&O builds. Round 6 adds per-vehicle expected ranges: GSA MAS commercial 2.2-2.6x, DoD Secret non-SCIF 3.1-3.4x, DoE M&O 3.0-3.8x, etc. Flag for review only if MID falls outside its vehicle-specific band.
3. **Sheet 5 day-trip IF branch.** Without $\text{IF}(B7=0, \dots)$ on rows 8 and 10, a day trip (Nights=0) silently produces 150% M&IE instead of the 75% single-partial-day per FTR 301-11.101. This is a workbook-level silent-wrong-answer bug. Template now shows the IF branch explicitly.
4. **CALC+ discovery path added to JSON-paths block.** `aggregations.labor_category.buckets` with `key/doc_count` per bucket. Sits alongside the existing `wage_stats` and `histogram_percentiles` paths. Wave 3 S2 worker had to probe the raw response because this wasn't documented.
5. **Text-starting-with-equals promoted to top-level silent-wrong-answer trap.** Previously buried under Sheet 2 formatting notes. Any cell starting with =, +, -, or @ is parsed as a formula by Excel, applies to all sheets including Methodology prose.

6. **Cross-sheet DL hourly reference index called out explicitly.** Wave 3 S1 worker hit a \$16.9B fantasy total by indexing off row 4 (Aged Annual Wage) instead of row 5 (DL Hourly). Previously only the FBR index ($18+i*19$) was called out.

Round 7 patches queued (not shipped)

None block current ship state.

1. Named ranges instead of row-indexed cell references to eliminate row-drift fragility when title banners or extra preamble rows are added to Sheet 1.
2. Mandatory Step 8.5 "Run recalc and verify" rather than parenthetical inside Step 8.
3. Arithmetic consistency check before save (pick one LCAT, one scenario, verify $\text{FBR} \times \text{hours} \times \text{headcount}$ equals Summary row).
4. DoD cleared engineering worked example in Quick Start (exact Wave 3 S1 pattern).
5. Thin-corpus CALC+ labeling rule: below ~25 records, label as "indicative only, not statistical validation."
6. FFP-by-deliverable Structure B scaffolding expanded to match Structure A depth (CLIN column template, per-LCAT vs uniform formulas, worked example).
7. RSE rubric propagated from BLS skill into FFP Methodology guidance (< 5% defensible, 5-15% cite with range, > 15% directional only).
8. Adapt FFP workflow patterns into IGCE Builder CR and IGCE Builder LH/T&M skills (already done for 17 cross-cutting patches; Round 5/6 additions not yet ported).
9. SOW decomposition Workflow A+ structured edit gate (add LCAT / rename LCAT / remove LCAT / split) before Step 1.
10. Sheet 2 block size constant cell for future-proofing if block row count changes.

Independent grading methodology

The Wave 1 and Wave 2 testing records were produced under a consistent methodology:

- Scenarios and assertion matrices were committed in writing before any worker output was read
- The grader did not coach workers during runs
- Assertions were graded strict on literal wording; ambiguous assertions were noted and refined for the next wave (not retroactively reinterpreted)

- Methodology is auditable in the `igce-ffp-wave1-runbook.md` and `igce-ffp-wave2-runbook.md` source files
- All findings come from direct observation of worker output and produced workbooks, not inference from memory of prior sessions
- Downstream skill patches (BLS Rounds 2 and 3, CALC+ Round 2) shipped before IGCE FFP Wave 2 so the substrate was validated in the Opus retest

Testing record prepared April 2026 by James Jenrette / 1102tools. Independent grading methodology. MIT licensed. Source: github.com/1102tools/federal-contracting-skills.