

**ccPlanning**

ccplanning.net

WORKFORCE PLANNING MASTERCLASS

# Quality Assurance Masterclass

Scoring what matters, calibrating it fairly, and turning QA into better service

**John Casey**

Workforce Planning Expert · ccPlanning

A masterclass in the ccPlanning white-paper series · read the rest at [ccplanning.net](http://ccplanning.net)

**Contents**

- 1. Quality is the point — and a planning concern..... 3
- 2. Score the substance, not the script..... 3
- 3. Calibration — make the scorers agree..... 4
- 4. Sample honestly..... 5
- 5. Coach from it — close the loop..... 5
- 6. AI-led versus human QA..... 5
- 7. A QA programme that improves service..... 6

## 1. Quality is the point — and a planning concern

*Service level and AHT measure the machine; quality tries to measure whether you actually helped. It is the closest thing to what the contact centre exists for — and it is far more a planning concern than planners usually admit.*

The link is direct. A poor-quality contact that fails to resolve the issue comes back as a repeat — a brand-new contact in next week's volume forecast. Rushed calls driven by an AHT target, agents pushed past sustainable occupancy, a skills gap that routes the wrong person to the wrong query: all of these show up first as a quality problem and then as extra workload. So QA is not a side function the planner can ignore; the quality programme and the staffing plan are two ends of the same system.

This masterclass is about running QA so that it actually improves service rather than generating tidy scores nobody acts on: what to score, how to calibrate it, how to sample honestly, how to coach from it, where AI fits, and how to wire the whole thing into an operating rhythm.

*A repeat contact is a brand-new contact in next week's forecast. Quality and capacity are two ends of the same system.*

## 2. Score the substance, not the script

The single biggest QA failure is a form that rewards compliance over outcomes. A checklist that scores whether the agent said the exact greeting, ticked every script step and stayed under a word count measures conformity, not quality — and it can actively punish the agent who broke from the script to do the right thing for a customer.

### Score the substance, not the script



*A form that rewards saying the script can punish actually helping.*

*Build the form around outcomes and behaviours that matter: resolution, accuracy, empathy, the right judgement — not script-ticking.*

Build the form around the things that genuinely make a contact good: did it resolve the issue, was the information accurate and compliant, did the agent listen and show empathy, did they make a sensible judgement in the grey areas, and did they set the right expectation? Weight

those, keep the form short enough that scorers can apply it consistently, and be ruthless about cutting items that measure the producer's convenience rather than the customer's experience.

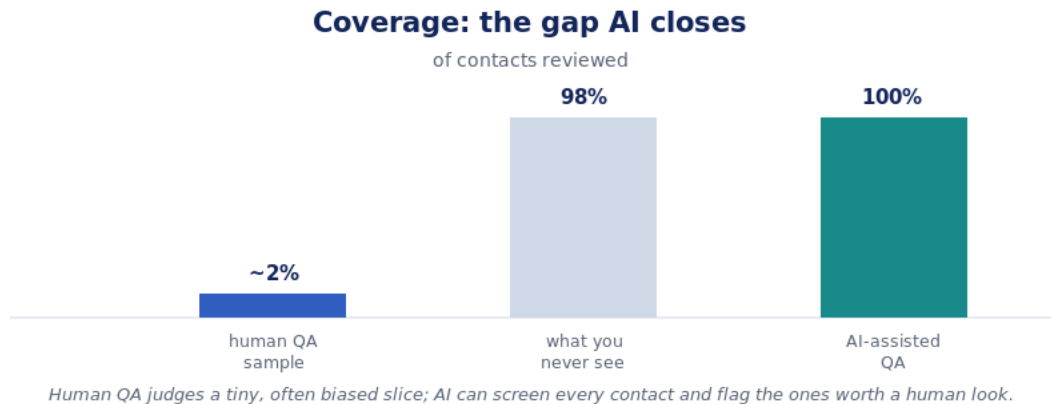
### **3. Calibration — make the scorers agree**

A QA score is only as good as the consistency of the people giving it. If two evaluators score the same call five points apart, the number means nothing — and agents quickly learn that their result depends on who happened to assess them, which destroys trust in the whole programme. Calibration is the discipline that fixes this: evaluators regularly score the same contacts independently, then compare and talk through the differences until they converge.

Done well, calibration is not a one-off; it is a standing rhythm that keeps the rubric interpreted the same way over time and surfaces the genuinely ambiguous items in the form (which are usually the ones worth rewording). Without it, your quality trend is measuring evaluator drift as much as agent performance.

## 4. Sample honestly

Most QA runs on a tiny, skewed slice of reality — a handful of evaluations per agent per month — and then that slice gets treated as truth. A three-point move on ten calls is noise, not a trend, and pinning a difficult conversation on it is managing randomness. Always ask “out of how many?” before you act on a QA number.



*Human QA sees a sliver; the 98% you never review is where most of the risk and the learning actually live.*

Watch for bias in what gets sampled, too: if evaluators cherry-pick or only review escalations, the picture skews. The honest approach is a representative sample, an explicit acknowledgement of the small-number uncertainty, and a refusal to build individual league tables off noise. Read QA trends in aggregate, where the sample is large enough to mean something, and use the individual scores as a coaching prompt rather than a verdict.

## 5. Coach from it — close the loop

A QA score that does not change behaviour is decoration. The point of the programme is the coaching conversation it enables: specific, kind, and aimed at one or two things the agent can actually improve, anchored in a real contact they can hear or read back. Scores filed in a spreadsheet nobody discusses improve nothing.

Coaching also feeds back up the system. A pattern of the same miss across many agents is rarely an agent problem — it is a training gap, a confusing process, a bad system, or an impossible target. The best QA programmes route those patterns to the people who can fix the cause, so the score becomes a diagnostic for the operation, not just a verdict on individuals.

## 6. AI-led versus human QA

AI quality tools can now screen every contact, which transforms the coverage problem: instead of a 2% sample, you can flag risk, sentiment and compliance issues across 100% of interactions, and surface the contacts that actually warrant a human look. That is a genuine advance, and it is where AI clearly wins — coverage, consistency and trend-spotting at a scale no human team can match.

Where humans still win is judgement: the grey areas, the nuance of an empathetic call that broke the script for good reason, and above all the coaching conversation that changes behaviour. The strongest programmes are not AI-versus-human but AI-and-human — automation does the screening and the breadth, people do the judgement and the development. Used that way, AI frees QA analysts from sampling drudgery to spend their time where humans add the most.

*AI gives you coverage and consistency; humans give you judgement and coaching. The best QA programmes use both for what each does best.*

## 7. A QA programme that improves service

Pulling it together into an operating rhythm — a checklist for a programme that actually moves the needle:

- **Score outcomes, not the script** — a short form built around resolution, accuracy, empathy and judgement.
- **Calibrate on a standing rhythm** so the number means the same thing whoever gives it.
- **Sample representatively** and respect the small-number uncertainty — no league tables off noise.
- **Coach from every evaluation** — specific, kind, anchored in a real contact, aimed at one or two things.
- **Route patterns to the cause** — training, process, systems or targets, not just the agent.
- **Use AI for coverage, humans for judgement** — screen everything, develop people.
- **Close the loop to planning** — feed repeat-driving quality issues back into the forecast and the staffing plan.

Run QA that way and it stops being an audit that agents dread and becomes what it should be: the feedback system that makes the service, and the plan behind it, steadily better.

---

### About the author

**John Casey** has spent more than 30 years in contact-centre workforce planning, including roles as a workforce-planning manager and operations director, and now writes and teaches at [ccplanning.net](https://ccplanning.net). The views in this paper are his own. It represents general professional experience and is not affiliated with, nor does it represent, any current or former employer.

[ccplanning.net](https://ccplanning.net) · Free guides, calculators and the rest of the white-paper series.