

Every difference, claimed

The theory behind the Tutors release harness

The one question

Is every observable difference between production and the candidate **claimed**?

- Not "is the candidate identical?" A release *should* differ.
- The harness proves that **every** difference is one somebody intended, and that nothing else moved.
- Same scripted traffic through both, noise normalised, everything observable diffed.

Main is the release candidate

- Trunk-based: every merge leaves `main` deployable. It need not ship, but it *could*.
- Students only ever see a **release**. `main` is staging.
- So the useful question every night is: *what would a release cut from `main` today look like beside production?*
- That nightly forecast is **Main to RC**. A real candidate (`vX.Y.Z-rc.N`) gets the same judgement as the gate.

A judge that lives apart

- A separate repository: the PR being judged cannot quietly weaken its judge.
- It compares **images, never checkouts**, so it can compare any two tags, even ones it did not build.
- Every image is pulled from `quay.io/tutors-sdk` and its **cosign signature verified**. An unsigned image is never judged.
- Every report says where each side's images came from.

Lean underneath

Idea	In the harness
Jidoka , stop the line	The Gate stops a release on an unclaimed difference
Visual management	One score, broken down, read the same way by everyone
Standard work	Twelve SOP steps, each with an owner and a done - when
Gemba , go and see	The reviewer's glance: at most seven places to look
Kaizen	Every stop ends in a countermeasure to the system, never blame

Explain, don't just count

"402 unclaimed" says nothing on its own. The harness has to say:

X differences. A, B and C contributed. That leaves **Y unexplained.**

- Count **causes, not hunks**: a dependency bump that adds 169 SBOM packages is one cause.
- Attribute each cause to a **claim** or to the **PR** that introduced it.
- What is left is the work: the unexplained gap.

The Gate wins

1. **Gate**: PASS, WARN or FAIL. It alone decides the exit code.
2. **Release Confidence Score**: 0 to 100 with a band. Advisory.
3. **Glance**: at most seven ranked places for the reviewer to look.

A FAIL is a FAIL at a score of 99. Nothing advisory (the score, the A3, the timing charts) ever changes a verdict.

Takeaways

- The harness asks one binary question and keeps it binary.
- It earns its right to fail, every night, by proving it is quiet against itself.
- Its job is to make a release decision **informed**: what changed, who changed it, and what nobody can explain yet.

Reports: tutors-sdk.github.io/tutors-release-harness