

# Inside the Evidence: What We Know, What We Don't, and Why We're Still Testing

*We could have hidden this page. Most organisations do. Here is every gap in our own evidence base, published on purpose — and the exact system we built to catch our own mistakes before they became public ones.*

Gemma Gardner — Founder, AURIS Institute™ · Public Research Briefing · July 2026

Most organisations publish their findings. Very few publish their gaps. We think that's backwards — if you're asking people to trust evidence-informed work, the honest version includes what the evidence doesn't yet support, not just what it does. This piece is that honest version, in full, with names and reference numbers attached rather than vague reassurance.

## 1. The System Behind Every Claim We Make

Everything AURIS publishes is checked against a single internal control document: the AURIS Master Evidence Repository. It currently holds close to 100 evidence rows across military transition, blue-light services, addiction recovery, elite sport, identity psychology, behavioural science and implementation science, each tagged with a source type, an evidence-strength rating, and a verification status. The rule governing all of it is deliberately blunt: if we cannot source it, we do not say it.

### 1.1 How a claim earns the right to be published

No claim goes into a public AURIS document, a funding application, or a parliamentary submission without first passing through the repository's claim-control process. Each substantive claim is logged in a Master Claim Register, given an approved, deliberately hedged wording, and checked against its current evidence strength — from Hypothesis, through Emerging and Moderate, up to Strong. The register also records unsafe wording to actively avoid: for example, the safe form of our central policy claim is "employment is important, but may not fully capture identity, belonging, purpose or long-term adaptation" — the unsafe form we explicitly rule out is anything resembling "employment does not matter" or "the CTP has failed." That distinction is enforced deliberately, not left to whoever happens to be writing on a given day.

### 1.2 Colour-coded, not just described

Every source in the repository carries a colour status — RED, AMBER or GREEN — alongside its narrative verification notes. GREEN means verified against a primary, published source. AMBER means a source exists but carries meaningful caveats that must travel with any citation of it. RED means not currently admissible in any public-facing document at all. This is not a formality: several sources that looked promising in early research passes were downgraded once we checked them properly, and we'll walk through a real example of that later in this piece.

## 2. Where the Evidence Is Strong

---

A meaningful share of the foundational claims behind AURIS are now verified against primary, published sources — not media summaries or second-hand citations. That includes core military-transition literature such as Grimell's ten-year identity study [1], official government transition statistics [2], and a 2024 study of the identities of spouses and partners of UK Armed Forces veterans [3] — the only population currently in our evidence base that examines the ripple effect of identity disruption beyond the individual who experienced it directly. It also includes the foundational psychological theory underpinning the AURIS Framework's five pillars: narrative identity theory [4], role-exit theory [5], self-efficacy theory [6] and salutogenesis [7]. Three of our own papers — the proof-of-concept research base, the Identity Performance Psychology framework, and the full white paper — are published openly on Zenodo with permanent DOIs, so the underlying work is checkable by anyone, not just cited by us.

## 3. Where the Evidence Is Genuinely Still Open

---

This is the part most organisations leave out, so here it is in full, not softened. Five specific, named gaps sit in our repository right now, each with its own priority rating and its own plan for closing it.

### Gap 1 — We don't yet have identity-specific evidence for every sector we work with

Policing is the clearest example. There's solid evidence on police wellbeing from the National Police Wellbeing Survey, run by Oscar Kilo, Durham University and the College of Policing [8] — but no peer-reviewed study yet located that examines identity disruption in policing specifically, as distinct from general occupational stress. A related police source, on sickness absence, remains RED in our repository — not currently admissible — while the wellbeing survey itself is GREEN and usable, but only for wellbeing context, not for an identity-mechanism claim. Until dedicated identity-specific research exists, we treat police-sector identity claims as a hypothesis worth testing, not an established finding, and we say so every time the subject comes up.

### Gap 2 — Our interview research is still early

The Identity After the Uniform programme has a target of 108 in-depth interviews. As of this piece, 22 have been recorded, transcribed and fully coded — not partially coded, not awaiting analysis, but fully coded and logged in our Interview Coding Matrix with no outstanding gaps in that subset. The remaining gap is not "finish coding the 22" — it's expanding the sample toward the full 108. That is real, rigorously coded qualitative evidence, but it is 22 of 108, and we disclose that denominator every single time we reference it, because early-stage themes drawn from a fifth of a planned sample can shift as the sample grows.

### Gap 3 — We caught our own inconsistency, and we're leaving the record of it visible

A recent internal audit flagged that one secondary source on NHS nursing burnout had been cited with three different, mutually inconsistent headline statistics across different internal documents — 2%, 2.75%, and a figure of 75.2% that appears to have been misread or mistyped at some point and then propagated. A related secondary summary of the same source also incorrectly claimed a peer-reviewed status in *Frontiers in Psychiatry* that it did not actually have. We have stopped citing any specific figure from this source until the discrepancy is resolved at the primary-source level, and the correction is tracked as its own numbered

entry — CR-005 — in our internal Contradiction Register, rather than being quietly edited away and forgotten. We are naming this publicly because a research programme that only shows you the mistakes it has already fixed, and never one it is still working through, is not being fully honest about how evidence work actually happens.

#### Gap 4 — Pilot delivery is not proof, and we won't pretend otherwise

Running a pilot programme demonstrates that a model can be delivered. It does not, by itself, demonstrate that the model works. Our evaluation framework requires pilot evidence to be assessed against implementation-science measures — acceptability, feasibility, fidelity to the model as designed, adoption, and participant-reported outcomes — before any effectiveness claim is made. Those evaluation protocols are still being finalised for our live pilots: ethics approval and follow-up arrangements are recorded as "to confirm" for all three current pilots in our own pilot register. We're not there yet, and we're telling you that rather than skipping ahead to a conclusion the evidence hasn't reached.

#### Gap 5 — We don't yet have direct evidence for coaching over therapy, specifically

AURIS is explicitly positioned as identity coaching, not clinical or therapeutic support — a boundary that matters both ethically and practically. We do have supporting evidence that non-clinical, goal-directed coaching is a recognised, evidence-based approach for populations pursuing personal change [9], a systematic review showing coaching is associated with promising, if still limited, behaviour-change outcomes generally [10], and a UK evaluation showing veteran-led, non-clinical peer support is well received alongside clinical care [11]. What we don't have — and haven't found anywhere in the literature despite specifically looking — is a study directly comparing coaching against therapy for identity transition specifically. That gap is logged, and it stays open until either the evidence turns up or we run the comparison ourselves through our own pilot evaluations.

*Evidence-integrity note: source [11] carries a moderate conflict-of-interest flag, as the study's author is employed by the organisation whose service was being evaluated, and the survey's response rate was approximately 19% from a self-selected sample. We cite it with that caveat attached in every use, not as unqualified proof.*

## 4. What the Repository Actually Contains, Sheet by Sheet

It may help to see the shape of the system itself, rather than just its outputs. The repository is not a single list — it is close to forty linked working sheets, each doing a specific job in the chain from raw source to publishable sentence.

### 6.1 The core registers

An Evidence Register holds every individual source — author, year, title, study type, population, key finding, colour status and exact citation — one row per source. A Claim Register sits above it, holding every substantive claim we might want to make, its approved safe wording, the specific evidence IDs that support it, and a deliberately named "unsafe wording to avoid" column for each one. A Gap Matrix tracks every unresolved evidence need by priority, owner action and current status — the five gaps described above are drawn directly from it, not summarised from memory. A Contradiction Register logs exactly the kind of internal inconsistency described above with Gap 3, each with its own reference number, so a correction has a permanent paper trail rather than disappearing into a silent edit.

---

## 6.2 The quality and traceability layer

A Risk-of-Bias Register and an Evidence Quality Scoring sheet assess each source on its own methodological merits — sample size, study design, conflicts of interest, peer-review status — and assign a numeric quality score out of ten, not just a colour. A Claim Traceability sheet maps every sentence in our most formal outputs, including the Parliamentary Evidence Review, back to the specific register rows that support it, so no sentence can outrun its evidence by accident. A Research Questions Register — covered in full in the next piece in this series — holds the formal, tracked hypotheses we are actively testing through our pilot programmes, distinct from claims about the published literature.

## 6.3 Why this level of structure, for an organisation this size

It would be reasonable to ask whether a founder-led organisation needs quite this much internal architecture. Our answer is that the architecture is precisely what makes a founder-led organisation's public claims trustworthy without an external institution vouching for them yet. Independent academic review of our work, through James Alder, is ongoing but has not concluded. Until it has, the repository's internal discipline — the fact that every claim has a traceable, checkable origin, and that gaps are logged rather than smoothed over — is the primary safeguard against overclaiming. We would rather over-invest in that structure now than have to retract a public claim later.

---

## 5. Quality Scoring in Practice — Not All GREEN Sources Are Equal

A source being GREEN — verified against a primary source — tells you it can be cited safely. It does not tell you how much weight the finding should carry. Our quality scoring sheet rates each GREEN source out of ten on its own methodological strength, and the range across our repository is genuinely wide. Grimell's ten-year longitudinal study, for instance, scores highly for its unusual duration and depth, while being explicitly flagged as a small sample from a single national context. The government's own CTP employment statistics score highly as official, large-sample administrative data, but are explicitly scoped as an employment measure only — not evidence about wellbeing or identity. Ingle's doctoral dissertation scores at the lower end — directional, useful, but unpublished and not yet peer-reviewed — and every citation of it in our public material carries that caveat rather than presenting it with the same weight as a longitudinal peer-reviewed study. This kind of differentiated, source-by-source honesty is, we think, rarer than it should be in research-adjacent public communication generally.

---

## 6. What "Safe for the Public" and "Safe for Parliament" Actually Mean

Every source in the repository carries two separate public-use flags, not one: whether it is safe to cite on the website, and whether it is safe to cite in a parliamentary or policy submission. These are not always the same answer. A source can be perfectly safe for general public reading — accurately described, properly attributed — while still requiring an evidence-level caveat before it goes anywhere near a parliamentary document, where the bar for precision is deliberately higher. This two-tier system is why the same underlying evidence base can support both an accessible piece like this one and a formally structured, 60-claim-register-backed Parliamentary Evidence Review, without either document overstating what the other has established.

---

## 7. How We Actually Check a Source — A Worked Example

---

It's one thing to say a claim is "verified." It's more useful to show what that actually involves. Take Grimell's ten-year identity study, one of the most important sources in our entire repository. Before it was marked GREEN, we confirmed: the exact DOI resolves to the published article itself, not a summary of it; the sample size and country (nineteen Swedish service members, annual interviews 2013–2016, a follow-up cycle in 2023) are stated accurately wherever we cite it; the study's own stated limitations — a small sample, a single national context — travel with the citation every time, not just in the footnotes; and the finding we draw from it is the finding the study actually supports, not a stronger version of it. That last check matters most. It would be easy to cite Grimell in support of a much bigger claim than the study actually makes. Our claim-control process exists specifically to stop that kind of drift before it happens, not to catch it afterwards.

---

## 8. Why We're Showing You This Instead of Hiding It

---

None of this is a confession of weakness. It's the discipline that makes the parts of our evidence base that are solid worth trusting. A research programme that never shows you an open gap either hasn't looked hard enough, or isn't telling you what it found. We'd rather be the organisation that shows its working — including the parts that aren't finished — because the population we're building this for has usually had enough of being oversold to.

The next piece in this series shows what happens when this evidence base meets real delivery — the specific questions our current pilot programmes are designed to answer, and why "the pilot is running" is a very different claim from "the pilot has proven something."

---

## References

---

- [1] Grimell, J. (2024). *Frontiers in Sociology*. <https://doi.org/10.3389/fsoc.2024.1406710>
- [2] Ministry of Defence (Defence Statistics) (2026). *Career Transition Partnership Annual Statistics 2024/25*. gov.uk
- [3] Spikol, E., McGlinchey, E., Fear, N., Armour, C., & Gribble, R. (2024). *Journal of Military, Veteran and Family Health*. <https://doi.org/10.3138/jmvfh-2023-0068>
- [4] McAdams, D. P. (2001). *Review of General Psychology*. <https://doi.org/10.1037/0033-2909.100.1.143>
- [5] Ebaugh, H. R. F. (1988). *Becoming an Ex: The Process of Role Exit*. University of Chicago Press.
- [6] Bandura, A. (1977). *Psychological Review*. <https://doi.org/10.1037/0033-295X.84.2.191>
- [7] Antonovsky, A. (1979/1987). *Health, Stress and Coping / Unraveling the Mystery of Health*.
- [8] National Police Wellbeing Service (Oscar Kilo), Durham University, College of Policing (2023/24). *National Police Wellbeing Survey*. [oscar.kilo.org.uk](https://oscar.kilo.org.uk)
- [9] Jarosz, J. (2016). *International Journal of Evidence Based Coaching and Mentoring*. <https://doi.org/10.24384/000102>
- [10] Ammentorp, J. et al. (2013). *BMC Health Services Research*. <https://doi.org/10.1186/1472-6963-13-428>
- [11] Aitken, D. & Kinane, C. (2024). *Journal of Veterans Studies*. <https://doi.org/10.21061/jvs.v10i1.480> — cited with response-rate and conflict-of-interest caveats.