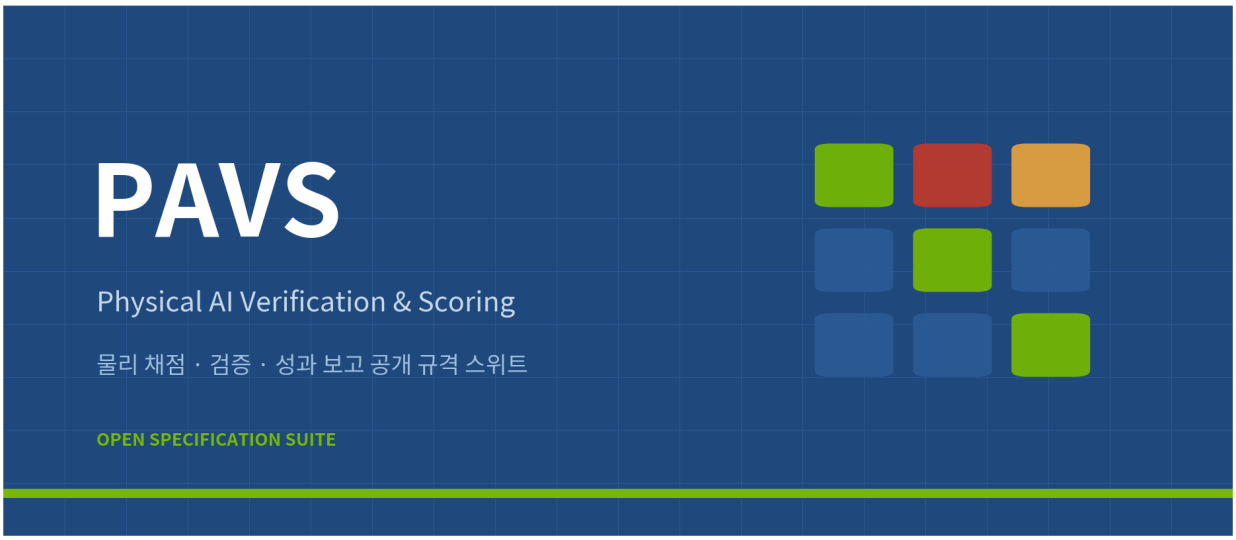



# Part 2 — 사례집 (해설·informative)

**발행본 v1.3 · 2026. 10. 29.**

발행: ㈜이머시브캐스트 · 제안 단계의 공개 규격(공식 표준화 기구의 표준이 아님)

immersivecast.com/pavs

# 목 차

**읽는 법 (informative)**

**A. 문항 — 측정 대상의 정의**

**B. 계측 — 값의 출처**

**C. 판정 — 합격 기준**

**D. 검증 — 채점기를 채점한다**

**E. 보고 — 결과의 기술 방식**

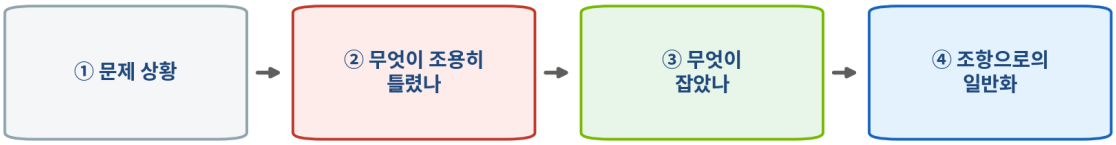
**맺음 (informative)**

**Physical AI Verification & Scoring — 조항이 태어난 사건들**

> **문서 지위**: 발행본 v1.3 · 2026-10-29 · 직전 발행 v1.2 2026-10-19 · 최초 공개 v1.0 2026-09-09 · 공개 위치 immersivecast.com/pavs · 대상 규격 = PAVS Part 1 v1.5(단언 72 · ★핵심 8 · v1.2~v1.4 와 요구 동일). **이 사례집(해설)은 Part 1 본체 및 부속서에 규정한 사항과 이에 관련된 사항을 설명하는 것으로, 규격의 일부는 아니다.** 이 문서는 전체가 informative(설명적)이며, 사례 서술 중의 요구 표현은 Part 1 조항의 요약일 뿐 이 문서가 요구사항을 창설하지 아니한다. **판 이력**: v1.0(2026-09-09, 33 건) → v1.1(2026-10-02, 38 건 · 신규 5 · 조항 대응 정정) → v1.2(2026-10-19, 39 건 · 신규 1 · 사례 D14 · D11 후속 비고 · 표기 정정) → v1.3(2026-10-20, 39 건 · 정오 정정 — PAVS 자기 지칭 「표준」 → 「규격」 · 다른 판 인용 삭제 · C3 · E2 제목 낱말 · 「제조사」 → 「제조자」 · 관련 조항 약칭 2 건 · 서술 규약 이름표 · 한글 표현 규범 용어집 일반어 교체: 「파지」 → 「집기」 · 「등재」 → 「등록 · 기록」 · 「갭」 → 「결손 · 간격」 · 「임계」 → 「기준값」 · 「지령」 → 「명령」). **위상**: 항공 안전 규정이 사고 조사에서 나왔듯, PAVS 의 조항들은 시뮬레이션 기반 로봇 검증을 실제 운영하며 겪은 사건에서 일반화되었다. 이 문서는 그 기록이다. 사례는 성과의 나열이 아니다 — 다수가 "검증 체계 자신의 결함을 검증 체계의 장치가 잡아낸" 기록이며, 자기 자신에게 동일한 기준을 적용하는 것이 이 규격의 목표 상태다. **서술 규약**: 모든 사례는 4 단 구조 — ①문제 상황 ②무엇이 조용히 틀렸나 ③무엇이 잡았나 ④조항으로의 일반화(사례 안 이름표는 「일반화」). 각 사례 머리에 관련 조항과 사건 시기(연월)를 적었다. 시스템·기종·식별자는 익명화했고, 수치는 사건의 본질에 필수적인 측정값만 사건 시기의 값으로 남겼다.

## 읽는 법 (informative)

**사례 서술의 4단 구조 — '에러 없이, 그럴듯하게, 조용히 틀린다'의 기록**




이 사례들의 공통 형태는 **'오류 표시 없이, 타당해 보이는 형태로, 조용히 틀린다'**이다. 시스템은 정상 종료되었고, 로그에 오류가 없었으며, 수치도 산출되었다. 그 수치가 거짓 합격 또는 거짓 불합격이었다. 규격이 정한 채점·검증·보고의 요구사항이 존재하는 이유는 바로 이 형태의 오류가 사람의 주의력으로는 잡히지 않기 때문이다.

사례는 Part 1 의 요구사항 6 군 중 다섯 사슬 군을 따라 배열했다: 문항(A) → 계측(B) → 판정(C) → 검증(D) → 보고(E). 운영(OPS) 군의 사례는 개정판 수록 후보로 관리한다. **사례 D6 은 이 규격의 참조 구현이 자기 자신을 점검하다가 발견한 사건이며, 그 결과 조항 하나가 신설되었다** — 규격이 사건에서 태어난다는 이 문서의 전제가 스위트 안에서 한 번 더 작동한 기록이다.

## A. 문항 — 측정 대상의 정의

### 사례 A1. 속도 0.8 이 힘 5.0 보다 작아서 합격했다

**관련 조항**: **PAVS-ITM-03**(단위 선언), **PAVS-ITM-04**(지표의 전 시험 종류 계측 보장)

**사건 시기**: 2026-08

**문제 상황.** 협동 안전 문항 "접촉 감지 기준값 ≤ 5.0 N"이 합격 판정을 내고 있었다.

**무엇이 조용히 틀렸나.** 그 문항에 공급된 값은 힘(N)이 아니라 속도 0.8 m/s 였다. 숫자만 보면 0.8 < 5.0 이므로 합격이다. 전수 조사 결과 17 개 문항이 자기와 다른 물리량으로 채점되고 있었다 — 충돌까지의 시간(초)이 접촉력(N) 문항을, 정지 횟수가 응답 시간 문항을 채점했다. 근본 원인은 조항 번호를 의미로 쓴 것이다: 참조 표준의 같은 절 번호가 판본에 따라 다른 물리량을 가리킨다.

**무엇이 잡았나.** 각 계측이 "측정 대상의 정의"를 단위로 선언하게 하고, 문항의 요구 단위와 다르면 채점 자체를 거부(미측정 처리 + 사유 기록)하도록 바꿨다. 이 수정이 판정을 관대하게 만들지 않았는지는 적대적 표본으로 별도 검증했다 — 틀린 판정의 철회는 "미측정"이 되지, 합격이 되지 않는다.

**일반화.** 계류(anchoring)는 조항 번호가 아니라 물리량 선언으로 한다. 단위가 맞지 않는 채점은 오류가 아니라 **없었던 시험**이다.

## 사례 A2. 문항의 명칭과 실제 계측 대상이 불일치했다

**관련 조항**: **PAVS-ITM-02**(단언 1:1), **PAVS-ITM-05**(변별력 실측)

**사건 시기**: 2026-08

**문제 상황.** "장애물 이격 유지" 문항이 오랫동안 안정적으로 합격을 내고 있었다.

**무엇이 조용히 틀렸나.** 판정 근거 경로에서 확인되었다. 실제 시험의 판정 근거 문자열에 남은 실측 키는 이격이 아니라 **총 이동거리**였다. 명칭과 실측 대상이 상이하면 문항은 자신이 측정하지 않는 항목을 보증하는 문서가 되며, 그 상태에서 통과 기록이 누적된다.

**무엇이 잡았나.** 문항-단언-검사의 1:1 대응을 감사하는 절차. 문항 이름이 아니라 실행되는 검사의 실체를 대조했다.

**일반화.** 문항의 실체는 명칭이 아니라 그것이 실행하는 검사다. 모든 문항은 실행 가능한 단언과 1:1 로 대응해야 하며, 그 대응 자체가 감사 대상이다.

## 사례 A3. 한 방향으로만 나오는 문항은 시험이 아닐 수 있다

**관련 조항**: **PAVS-ITM-05**(변별력), **PAVS-ITM-06**(반대 방향 대조군), **PAVS-ITM-07**(코퍼스)

**사건 시기**: 2026-08

**문제 상황.** 속도 제한 구역 준수 문항이 증명된 시험 이력 전부에서 위반만을 기록하고 있었다 — 준수 판정이 한 번도 나온 적이 없었다.

**무엇이 조용히 틀렸나.** 단일 방향만 산출하는 문항에는 두 가지 해석이 가능하다. 실제로 전건이 위반이었을 가능성과, 게이트가 반대 값을 산출할 수 없는 구조일 가능성이다. 후자라면 그 문항은 시험으로 기능하지 않으며, 두 상태는 외형으로 구별되지 않는다.

**무엇이 잡았나.** 반대 방향 대조군 — 원본 시나리오에서 자극 한 줄만 준수 방향으로 변경한 대응 시나리오 — 를 만들어 준수 판정이 실제로 산출되는지 확인했다. 설계 원칙이 대조군 파일 머리에 함께 남았다: 문항의 한계를 조정해 원하는 판정을 만드는 것은 검증이 아니라 조작이며, 대조군은 위반·준수·비율형 문항을 섞어 배치해 "값이 나왔을 때 어느 계측기가 살아있는지"까지 변별한다.

**일반화.** 규격의 실체는 문항 목록이 아니라 **시나리오 + 대조군 + 기대 판정의 실행 가능한 코퍼스**다. 단일 판정만 내 온 문항의 경계는 반대 방향 대조군으로만 닫는다.

## 사례 A4. 기준선에 제 3 자가 없으면, 같은 로봇이 합격도 불합격도 된다

**관련 조항**: **PAVS-ITM-08**(조건 있는 기준선)

**사건 시기**: 2026-08

**문제 상황.** 보도 주행 배송 과제에서, 보행자 밀도를 100 m 당 0.25 명으로 소폭 상향하자 성공이 5/5(100% · Wilson 95% 하한 56.6%)에서 0/5(0% · Wilson 95% 상한 43.4%)로 하락하였다(밀도마다 시드 5 개).

**무엇이 조용히 틀렸나.** 로봇의 성능이 저하된 것이 아니었다. 시간 기준선(SLA) 산식에 보행자 양보 소요 시간이 포함되지 않아, 무인 보도 기준으로 계산된 여유가 수 초에 불과하였다. 동일 산식 계열에서 이 유형의 누락(소요 시간 항목의 누락)은 이미 두 차례 반복된 상태였다. 구축 과정에서는 오류 표시 없이 타당해 보이는 수치를 산출하는 모델 결함도 연이어 검출되었다. 중심 거리로 접촉을 판정하여 정상 교행의 절반 가까이를 접촉으로 오판한 사례가 그것이다.

**무엇이 잡았나.** 기준선은 제 3 자 밀도 조건과 함께 정의되어야 한다는 원칙을 기록했다: 조건을 적지 않으면 같은 로봇이 같은 경로에서 합격도 되고 불합격도 된다.

**일반화.** 성능·안전 기준선은 그것이 성립하는 조건(주변 인간·환경·하중)과 함께 정의되어야 한다. 조건 없는 기준선은 재현 불가능한 판정이다. (이 사례의 보도 폭은 선언한 값 기반이며, 시험은 1 차원 운동학 모델에서 수행되었다.)

## 사례 A5. 그 시험 계열에서 원리적으로 잴 수 없는 문항이 걸려 있었다

**관련 조항**: **PAVS-ITM-04**(지표의 전 시험 종류 계측 보장), **PAVS-VDT-01**(3 상 판정 — 필수 문항의 계측 부재 처리)

**사건 시기**: 2026-08

**문제 상황.** 민수 항공 시나리오군(기반시설 점검·수색구조·도시 순찰 등)이 채점에서 일제히 불합격으로 나왔다.

**무엇이 조용히 틀렸나.** 해당 시나리오에 '충돌 없음' 문항이 걸려 있었으나, 이 계열의 비행체는 운동학 모델로 구동되어 **충돌량이 원리적으로 산출되지 않는다**. 27 개 시험을 전수 확인한 결과 실측 건수는 0 이었다. 판정 계통은 미측정을 합격으로 접지 않으므로 불합격을 냈고, **그 판정 자체는 옳았다.** 결함은 판정 계통이 아니라 그 계열에 적용될 수 없는 문항이 걸려 있었다는 데 있었다.

**무엇이 잡았나.** 미측정을 합격으로 접지 않는 판정 규칙. 이 규칙이 없었다면 산출되지 않은 값이 '위반 0'으로 읽혀 **거짓 합격**이 되었을 것이다. 조치는 기준값을 낮추거나 값을 채우는 것이 아니라 **해당 계열에서 그 문항을 제거**하는 것이었다. Part 1 v1.2 의 **PAVS-VDT-01** 은 이 처리를 명문화하였다 — 필수 문항의 계측 부재는 '계측 부재' 사유의 불합격으로 계상하고, 미측정 건수는 별도로 보존한다.

**일반화.** 문항은 그것이 적용되는 시험 계열에서 계측이 성립함을 전제한다. 계측이 성립하지 않는 문항은 그 계열에서 문항이 아니며, 불합격으로 남겨 두는 것도 합격으로 접는 것도 옳지 않다 — **문항 목록에서 빼고 그 사실을 기록하는 것**이 옳다. 정직한 판정 규칙은 이런 경우 '불합격'으로 신호를 보내며, 그 신호를 판정 계통의 결함으로 오인하면 규칙을 완화하는 방향으로 잘못 수정하게 된다.

## 사례 A6. 세 안전 조항이 늘 합격이었다 — 로봇이 안전해서가 아니었다

**관련 조항**: **PAVS-ITM-05**(변별력 실측 분류), **PAVS-ITM-06**(반대 방향 대조군)

**사건 시기**: 2026-08

**문제 상황.** 사람 이격·개인공간·대인 충돌에 관한 안전 조항 세 건이 이동 로봇 계열에서 한 번도 불합격을 낸 적이 없었다.

**무엇이 조용히 틀렸나.** 변별력을 확인하려고 **사람 10 명을 경로 위에 벽처럼 세운 대조군**을 만들었다. 결과는 개인공간 침범 0 건·대인 충돌 0 건이었으나, 함께 기록된 다른 값이 이유를 말했다 — 로봇의 총 이동거리가 2.04 m 였고, 사람과의 최소 거리는 3.01 m 였다. 런타임 설정이 **사람 3.5 m 앞에서 정지**하도록 되어 있었다. 개인공간 기준값은 1.2 m 이므로 **원리적으로 침범될 수 없다.** 세 조항의 상수 합격은 로봇의 미덕이 아니라 **런타임 보장**이었다.

**무엇이 잡았나.** 한 번도 불합격이 나지 않은 문항을 '미검증 경계'로 분류하고, 자극을 반대 방향으로 준 대조군으로만 닫도록 한 규율. 이 사례에서 그 문항은 **닫히지 않았다** — 시험이 성립하려면 정지 거리를 기준값보다 짧게 설정해야 하는데 그것은 안전하지 않은 구성이다. 따라서 '시험됨'이 아니라 '런타임 보장에 의한 상수 합격'으로 성과 문서에 기재하였다.

**일반화.** 항상 합격하는 문항은 두 가지 중 하나다 — 피시험체가 우수하거나, 그 상황에 **도달할 수 없거나**. 둘을 가르는 것은 대조군이며, 대조군 없이 누적된 합격은 능력의 증거가 아니다. 시험이 성립하지 않는 문항은 합격으로 계상하지 아니하고 그 사유와 함께 별도로 기재하여야 한다. **시나리오·대조군·기대 판정의 묶음이 곧 규격의 실물**이며, 조항 문안만으로는 이 구별이 성립하지 않는다.

### 사례 A7. 지표가 로봇이 아니라 시나리오를 재고 있었다

**관련 조항**: **PAVS-ITM-05**(변별력 실측 분류)

**사건 시기**: 2026-08

**문제 상황.** 작업공간 점유율 조항이 변별력을 갖는지 확인하기 위하여, 같은 로봇·같은 보행 경로에서 **구역 크기만** 15 m 에서 3 m 로 바꾸어 시험하였다.

**무엇이 조용히 틀렸나.** 판정이 0.618(불합격)에서 0.0(합격)으로 바뀌었다. 로봇의 거동은 동일하였다. 이 지표의 분모가 시험 전체의 관측 구간이었으므로, **값은 로봇의 속성이 아니라 시나리오의 속성**이었다. 조항은 변별력을 갖되 로봇을 변별하지 않았다.

**무엇이 잡았나.** 변별력 확인을 '판정의 갈림 여부'에서 멈추지 않고 **판정을 가르는 요인**까지 확인한 절차. 단일 변수만 바꾸는 통제 시험이 그 구별을 만들었다.

**일반화.** 판정이 갈린다는 사실만으로는 문항이 유효함을 보이지 못한다. 시나리오 설정을 바꾸었을 때만 갈리는 지표는 피시험체를 비교하는 근거가 되지 못하며, 서로 다른 시험 대상의 성적을 나란히 놓는 순간 **시나리오 차이가 능력 차이로 읽힌다.** 변별력의 분류에는 무엇을 변별하는가가 함께 기록되어야 한다.

## B. 계측 — 값의 출처

### 사례 B1. 최고속도가 실측이 아니라 "명령 × 1.35"였다

**관련 조항**: **PAVS-MEA-01**(출처 증명), **PAVS-MEA-07**(실효값 병기)

**사건 시기**: 2026-08

**문제 상황.** 이동 로봇 안전 채점에서 최고속도 지표가 여러 안전 조항의 판정에 쓰이고 있었다.

**무엇이 조용히 틀렸나.** 그 값은 로봇이 실제로 낸 속도가 아니라 주행 명령값에 계수를 곱한 것이었다. 실제 1.4 m/s 로 달린 로봇의 최고속도가 서류상 4.7 m/s 로 기록되어, 그 키에 묶인 안전 조항들에서 거짓 불합격이 산출될 수 있는 상태였다 — 그리고 존재하지 않는 과속이 위반 계수기를 실제로 켰다. 판별은 3 중 증거로 이뤄졌다: 값이 명령의 정확한 배수라는 것, 서로 다른 두 시험의 값이 소수점 끝자리까지 동일하다는 것(실측값에서는 성립할 수 없는 일

치), 총 이동거리와 물리적으로 모순된다는 것. 결정적 대조는 다른 시험 계열이었다 — 같은 이름의 지표가 그 계열에서는 실측이었다. **같은 지표 이름이 시험 계열마다 출처가 달랐던 것이다.**

**무엇이 잡았나.** 변위 기반의 실측 대조 계측을 신설해 명령 유래 값과 상시 병기하고, 값의 출처(실측/명령/기본값) 선언을 규칙화했다. 다만 명령 유래 값은 채점 입력에서 아직 제거되지 않고 실측값과 병기된 상태이며, 이 상태는 **PAVS-MEA-01**(비실측 값으로 채점하지 않음)의 미충족으로 기록되어 있다(이 판 기준).

**일반화.** 출처를 묻지 않는 채점은 양방향으로 틀린다 — 명령값을 실측으로 오인하면 거짓 불합격이, 부재를 0 으로 치환하면 거짓 합격이 난다. 같은 장치(출처 증명)가 둘 다 막는다.

## 사례 B2. 측정 주체 없는 0 이 '합격'으로 처리되었다

**관련 조항**: **PAVS-MEA-02**(결측≠0), **PAVS-VDT-02**(긍정 증거), **PAVS-REP-03**(커버리지 목록)

**사건 시기**: 2026-08

**문제 상황.** 충돌 횟수 0 이 안전 문항을 통과시키고 있었다.

**무엇이 조용히 틀렸나.** 그 0 은 측정 주체가 없는, 부재를 0 으로 치환한 값이었다. 같은 형태를 일제 조사하자 22 건이 나왔다: 시험 0 건인 항목이 "인증됨"(당시 표기)으로, 도구 사용 실적 0 건이 오류율 0 의 "적격"(당시 표기)으로, 관측되지 않은 충돌 시간이 "999 초 관측값"으로 보고되고 있었다 — 데이터 타입이 숫자를 강제하여 '없음'을 표현할 수단이 없는 경우에는 타입이 허위 값을 강제하였다. 백엔드에서 수정한 규칙을 표시 계층이 자체 공식으로 재계산하여 합격 표시를 복원한 사례도 있었다.

**무엇이 잡았나.** '측정값이 전혀 없는 입력에서 합격이 산출되면 위반'으로 판정하는 실행형 가드. 그 가드의 첫 실행이 수정 과정에서 생긴 새 결함 2 건(과교정 방향 포함)까지 잡았다.

**일반화.** 결측은 0 이 아니다 — 급정지를 하지 않았는데 0 초에 멈춘 것으로 기록되면 만점이 된다. 또한 판정 규칙을 단일 지점에 집약하지 않으면 수정 후에도 표시 계층에서 복원된다.

## 사례 B3. 고장 난 계측을 신뢰한 페루프가 개루프보다 나쁜 결과를 냈다

**관련 조항**: **PAVS-MEA-03**(liveness), **PAVS-MEA-04**(기지 자극 프로브)

**사건 시기**: 2026-07

**문제 상황.** 시각 보정 서보가 계측된 오차를 받아 집기 위치를 미세 조정하고 있었다. 다섯 번 시도해 다섯 번 실패했다.

**무엇이 조용히 틀렸나.** 보정 루프 자체가 들여쓰기 오류로 단 한 번도 실행된 적이 없었다 — 따라서 그 이전의 성공들은 이 보정의 공로가 아니었다. 루프를 복구하자 보정의 입력이 되는 계측값이 시도와 무관하게 **상수**임이 드러났고, 서보는 그것을 유효한 오차로 믿고 매번 같은 잘못된 보정을 주입해 전패했다.

**무엇이 잡았나.** 계측 입력의 살아있음(liveness) 검증 — 입력 상태가 변하는데 계측값이 불변이면 유효 신호가 아니라 계측 고장으로 처리한다. 동일 유형의 사건이 하나 더 있다. 카메라 좌표 관례 불일치로 렌더가 전면 무효였는데, 설정값 되읽기 검산은 같은 관례를 공유해 "정상"을 반환했다 — 판별은 알려진 표적(단색 물체)을 넣고 종단 출력의 픽셀을 세는 독립 프로브로만 가능했다.

**일반화.** 계측의 건전성은 자기일관 검산이 아니라 기지 자극의 종단 프로브로 확인한다. 고장 난 계측을 신뢰하는 페루프는 오류를 체계적으로 주입한다.

## 사례 B4. 채점기가 피시험체를 느리게 만들었다

**관련 조항**: **PAVS-MEA-08**(한계 하드코딩 금지)

**사건 시기**: 2026-02

**문제 상황.** 전 속도 범위로 훈련된 정책이 시험에서 설계 범위보다 낮은 속도와 불안정한 거동을 보였다.

**무엇이 조용히 틀렸나.** 채점 시스템 코드에 속도 상한이 상수로 내장되어 있었다. 피시험체는 자신의 설계 범위를 쓰지 못했고, 성적은 로봇의 능력이 아니라 채점기의 상수를 반영했다.

**무엇이 잡았나.** 모든 제한을 피시험체의 선언 사양(매니페스트)에서만 유래하게 하는 규칙.

**일반화.** 채점기는 피시험체를 왜곡해서는 안 된다. 시험 조건이 피시험체의 설계 조건과 다르면, 그 사실 자체가 결과에 명시되어야 한다.

## 사례 B5. 3.6 km 경로를 161 개 점으로 대표하였다

**관련 조항**: **PAVS-MEA-05**(프록시 지표 자격 명시)

**사건 시기**: 2026-08

**문제 상황.** 보도 주행 배송 과제에서 통로의 통과 가능 여부는 구간 끝점의 폭 한 값으로 판정되었다. 경로는 3.6 km 이고 폭 표본은 161 점이며, 표본 간격은 중앙 22.0 m·최대 34.6 m 였다.

**무엇이 조용히 틀렸나.** 끝점의 폭이 충분하면 구간 전체가 통과 가능으로 판정되었고, 표본과 표본 사이의 협소부는 판정에 들어오지 않았다. 한 구간에서 실제 측방 이격이 **−102 mm** 로 산출되었다 — 음수는 기하적으로 통과 불가를 뜻하나 판정은 통과였다. 끝점의 폭은 물리량이 아니라 **표본 간격이라는 이산화 조건의 함수**였으며, 그 자격이 선언되지 않은 채 구간의 대표값으로 쓰였다.

**무엇이 잡았나.** 판정의 입력을 끝점 폭에서 **구간 최소폭**으로 교체하였다. 같은 계통에서 공간 기준점을 로봇의 실제 호(弧)길이가 아니라 목표 웨이포인트로 두어 기준점이 한 틱에 13.9 m 도약한 결함도 함께 검출되었으며, 이는 추측이 아니라 계측 로그로 확인되었다.

**일반화.** 구간을 대표하는 표본은 그 구간에서 가장 불리한 값이어야 한다. 대표 범위가 선언되지 않은 지표는 프록시이며, 프록시의 자격을 적지 않으면 판정은 물리가 아니라 표본 간격의 함수가 된다. 이 사례에서 동일 결함이 선행 통과성 검사에도 존재하나, 수정 시 기존 14 개 시나리오의 판정이 함께 이동하므로 **미수정 결손으로 기록**하였다 — 결손을 기록하는 것과 결손을 닫는 것은 다른 사안이며, 기록 없이 닫지 않는 것이 더 나쁘다.

## 사례 B6. 같은 이름의 지표가 시험 계열마다 다른 뜻이었다

**관련 조항**: **PAVS-MEA-01**(출처 증명 · 비실측 채점 금지)

**사건 시기**: 2026-08

**문제 상황.** 사례 B1 에서 이동 로봇의 '최고속도' 지표가 실측이 아니라 순항 명령의 1.35 배임이 확인되었다. 같은 이름의 지표가 다리 로봇 계열에도 존재하였으므로, 전 계열 일괄 치환이 검토되었다.

**무엇이 조용히 틀렸나.** 다리 로봇 계열의 같은 지표는 **실측이었다**. 해당 기체의 값은 1.4985 m/s 였고 명령값은 2.2 m/s 이므로, 이동 로봇 계열에서 확인된 1.35 배 관계를 적용하면 2.97 m/s 가 되어야 하나 그렇지 않았다. **키의 이름이 같다는 것이 값의 출처가 같다는 근거가 되지 못하였다.** 일괄 치환을 수행하였다면 출처가 올바르던 계열의 값이 명령값으로 교체되어, 결함이 없던 쪽에 결함이 주입되었을 것이다.

**무엇이 잡았나.** 치환 이전에 계열별로 값의 출처를 개별 확인하는 절차. 판별은 세 가지를 함께 보았다 — 값과 명령의 정확한 배수 관계 여부, 실행 간 소수점 이하까지의 동일 여부, 총 이동거리와의 모순 여부.

**일반화.** 값의 출처는 지표의 이름이 아니라 그 값을 만든 계통에 속한다. 동일 명칭의 지표에 대한 일괄 수정은 출처를 계통별로 확인한 뒤에만 수행하여야 한다. 결함의 발견은 그 결함이 발견된 범위에만 적용되며, 범위를 확인하지 않은 확대 적용은 수정이 아니라 새로운 결함이다.

### 사례 B7. 반영값은 맞는데 효과는 없었다

**관련 조항**: **PAVS-MEA-07**(요청값·실효값·적용 병기), **PAVS-MEA-04**(기지 자극 종단 프로브), **PAVS-VRF-03**(실효값 확인)

**사건 시기**: 2026-10

**문제 상황.** 로봇 팔 위치 정확도 시험에 국제 표준이 요구하는 정격 하중 조건을 추가하였다. 시험 하중은 팔 끝 링크의 질량에 정격 가반 하중을 더하는 방식으로 적용하고, 적용 여부는 질량 되읽기로 확인하도록 하였다. 첫 시범 실행에서 되읽은 질량은 요청대로 +10 kg 이었다.

**무엇이 조용히 틀렸나.** 하중을 더한 뒤 같은 관절 목표로 정착시킨 팔 끝의 변위가 0.000 mm 였다. 10 kg 을 더한 팔이 전혀 처지지 않았다. 물리 엔진은 정지한 관절체를 휴면 상태로 두며, 질량 속성의 변경은 휴면을 깨우지 않는다. 설정값은 저장되었으나 물리에는 반영되지 않았다. 되읽기만으로 적용을 확인하였다면 이 실행은 '하중 적용됨'으로 기록된 채 무하중 측정값을 하중 조건의 결과로 내놓았을 것이다.

**무엇이 잡았나.** 효과 가드 — 하중 적용 전후 같은 관절 목표에서 팔 끝 처짐이 문턱(0.1 μm)을 넘지 않으면 실행을 종료하고 미측정으로 남기는 검사 — 가 첫 시범 실행에서 발화하였다. 원인을 추측으로 고치지 않고 판별 계측(깨우지 않은 상태의 변위, 관절 속도를 0 으로 다시 기록하여 깨운 뒤의 변위, 무하중 정상상태 관절 오차, 휴면 기준값)을 추가해 한 번 더 실행하였다. 깨우기 전 변위는 0, 깨운 뒤에는 약 1 mm 의 처짐이 나타나 휴면 가설이 확정되었고, 효과 판정의 기준을 깨운 뒤의 변위로 바꾸었다. 같은 기전이 기존 정확도 측정의 정착 구간을 왜곡했는지도 기록된 궤적으로 오프라인 점검하였으며, 추정 왜곡은 최악 정확도 대비 약 1% 이하였다. 결함은 결과 발행 전에 잡혔다.

**일반화.** 요청값·실효값 병기에서 실효값은 설정값의 되읽기가 아니라 **적용 결과의 되읽기**여야 한다. 설정값의 되읽기는 값이 저장되었음을 보일 뿐, 물리가 그 값대로 바뀌었음을 보이지 않는다 — 설정과 같은 경로로 하는 확인은 설정 경로의 결함을 공유한다(사례 B3 의 자기일관 검산과 같은 구조). 시험 조건의 실효는 그 조건이 만들어야 하는 물리 변화(하중이라면 처짐)를 종단에서 확인하여야 하며, 효과가 확인되지 않은 조건의 실행은 조건 적용 결과가 아니라 미측정이다. 비교 시험의 독립변수(사례 D3)에도 같은 규율이 적용된다.

## C. 판정 — 합격 기준

### 사례 C1. 기본값이 합격을 반환하고 있었다 — 그리고 401 건 전수 중 89%

**관련 조항**: **PAVS-VDT-02**(긍정 증거), **PAVS-VDT-07**(소급 감사)

**사건 시기**: 2026-05

**문제 상황.** 임무 성공 판정이 오랫동안 높은 합격률을 보이고 있었다.

**무엇이 조용히 틀렸나.** 판정 로직의 기본 경로가, 필수 조건(전 경유점 도달·대상물 실측 상승)이 미충족이어도 성공을 반환하고 있었다.

**무엇이 잡았나.** 합격을 "전 조건의 실측 충족"이라는 긍정 증거로만 산출하도록 수정했다 — 그리고 더 중요한 것: **결함 기간의 기존 합격 전수를 새 기준으로 소급 감사했다.** 결과는 결함 기간 합격 401 건 전수 중 89%가 거짓 합격(전

수 소급 감사 · 2026-05). 소급 감사 의무가 없었다면 이 수치는 존재하지 않았고, 거짓 합격은 실적으로 남았을 것이다.

**일반화.** 판정 기준의 결함 수정은 코드 수정으로 끝나지 않는다 — 그 기준이 산출했던 과거 판정의 소급 감사와 결과 공개까지가 수정이다.

## 사례 C2. 그리퍼 미부착 상태에서 '집기 성공'이 산출되었다

**관련 조항**: **PAVS-VDT-03**(물리 교차검증)

**사건 시기**: 2026-05 (유사 계열의 접촉력 판별 2026-07)

**문제 상황.** 픽앤플레이스 시험에서 집기 성공, 배치 성공, 임무 성공이 모두 계상되었고 화면 표시도 "파지 완료"였다.

**무엇이 조용히 틀렸나.** 그리퍼가 부착되지 않아 열림/닫힘 명령이 전부 무동작(no-op)이었다. 성공 지표는 명령의 발행을 세고 있었지, 물리적 사건을 세지 않았다.

**무엇이 잡았나.** 성공 지표를 물리 증거와 교차검증하는 규칙 — 대상물의 실측 상승, 접촉력, 영상. 유사 계열로, 물체를 손에 논리적으로 붙이는 방식(weld)의 "성공"은 접촉력 실측(정상 집기 약 20 N 대 weld 약 0 N)으로 판별되어 하드 불합격으로 강등되었다.

**일반화.** 성공 지표는 그것이 전제하는 물리적 사건의 독립 증거와 교차검증되어야 한다. 명령이 발행되었다는 것과 세계가 변했다는 것은 다른 명제다.

## 사례 C3. 대형 수상함이 침몰했으나 성적은 '합격'이었다

**관련 조항**: **PAVS-ITM-05**(변별력), **PAVS-REP-12**(안전 이상≠'정상') — 채점표의 커버리지

**사건 시기**: 2026-08

**문제 상황.** 해상 시나리오 재검증에서 대형 수상함이 시작 직후 등속으로 가라앉아 해저를 통과하여 하강하였다. 그럼에도 그 시험의 총평은 합격이었다.

**무엇이 조용히 틀렸나.** 채점표에 전복·침몰을 묻는 문항이 없었다. **채점표에 없는 실패는 조용히 통과한다.** 이 결함은 능력의 문제가 아니라 문항 커버리지의 문제이며, 정상 시험이 많을수록 오래 숨는다.

**무엇이 잡았나.** 개선된 채점 기준으로 과거 합격작 전량을 재시험하는 절차가 이 사건을 드러냈다. 재검증은 결과를 다음과 같이 구분한다. 판정 유지가 확인된 것, 판정 근거 자체가 없던 통과(미측정), 낡아 있던 판정, 그리고 새 기준이 드러낸 미감지 결함 — 계측기의 변화를 능력의 회귀로 오귀속하지 않도록 귀속을 분리 기록하면서다. 이 사건이 필요성을 드러낸 침몰·전복 문항은 이 판 기준(2026-10) 아직 판정 문항으로 구현되지 않았다 — 공백은 기록되어 있으나 닫히지 않았다.

**일반화.** 채점 체계는 주기적으로 자기 커버리지를 재시험해야 한다. "그때는 몰랐던 실패 모드"는 문항 신설과 소급 재검증으로만 잡힌다.

## 사례 C4. 같은 로봇, 같은 과제 — 제원만 바꾸면 판정이 정반대

**관련 조항**: **PAVS-VRF-05**(모델 진본성), **PAVS-VDT-08**(불가 명령 거부)

**사건 시기**: 2026-08

**문제 상황.** 동일한 로봇 팔의 동일한 과제를, 공개 모델 저장소의 제원과 제조자 확정 제원으로 각각 채점했다.

**무엇이 조용히 틀렸나.** 공개 제원은 손목 관절 한계가 실기보다 넓게 기재되어 있었다. 그 제원으로는 실기가 물리적으로 취할 수 없는 자세의 과제가 **합격**으로 산출됐다(거짓 합격). 같은 모델의 다른 축들은 실기보다 좁게 기재되어 가능한 자세를 잃게 만들었고, 실측에서는 그리퍼 구동 제원의 결함이 정격 이내 하중의 과제를 불합격시켰다(거짓 불합격). 제원 파일의 수치만 수정하는 것으로는 거동이 변경되지 않았다. 물리 엔진이 실제로 한계를 강제하는 지점은 별도에 있었다.

**무엇이 잡았나.** 제조자 확정 제원과의 대조 검증, 그리고 수행 불가능한 명령을 조용히 근사하지 않고 시험 자체를 거부하는 규칙(fail-loud) — 확정 제원 체계에서는 해당 과제가 "합격"이 아니라 "시험 거부"로 처리되었다.

**일반화.** 제 3 자를 채점하는 체계에서 모델 진본성은 채점의 전제다. 반입된 모델의 진본성을 먼저 검증하지 않으면 높은 성적도 아무것도 증명하지 못한다.

## 사례 C5. 로봇이 정지해 있는 동안의 통과가 로봇의 접촉으로 계상되었다

**관련 조항**: **PAVS-VRF-04**(피시험체 귀속 전 반증)

**사건 시기**: 2026-08

**문제 상황.** 보도 주행 배송 과제에서 보행자·자전거와의 접촉 건수를 계상하였다. 협소 구간에서 로봇이 정지한 채 대기하는 동안 보행자가 로봇 옆을 통과한 사건이 접촉으로 분류되었다.

**무엇이 조용히 틀렸나.** 통로 폭 762 mm 구간에서 접촉 5 건이 계상되었다. 로봇은 그 시점에 정지 상태였고 회피·감속·경로 변경 중 어느 것도 수행할 수 없었다. **사건은 실재하였으나 피시험체의 행위가 아니었으며**, 판정은 거짓 불합격이었다. 타자가 뒤에서 추월한 통과도 동일하게 분류되었다.

**무엇이 잡았나.** 분류 기준을 '그 시점에 행동할 수 있었던 주체'로 변경하였다. 피시험체가 행동할 수 없었던 통과는 접촉에서 제외하되 **별도 범주(보도 점유)로 계상**하여, 사건 자체는 성적에 남되 피시험체의 판정에는 들어가지 않도록 하였다.

**일반화.** 사건의 발생과 사건의 귀속은 다르다. 피시험체에 사건을 귀속하기 전에 피시험체가 그 사건에 대하여 행동할 수 있었는지를 확인하여야 한다. 행동 가능성이 없는 사건의 귀속은 거짓 불합격을 만들고, 그 사건을 삭제하는 것은 정보의 손실이므로 별도 범주로 보존하여야 한다. **PAVS-VRF-04** 는 귀속 이전의 반증 대상을 계측 계통의 건전성으로 한정하고 있으며, 이 사례는 그 대상이 **행동 가능성까지 확장되어야 함**을 시사한다 — 조항 확장 후보로 Part 1 12.6 의 차기 정식 개정 검토 대상에 포함하였다.

## 사례 C6. 코드 표 밖 몸체 접촉은 통과로 빠졌다

**관련 조항**: **PAVS-ITM-04**(지표의 전 시험 종류 계측 보장), **PAVS-VDT-02**(긍정 증거), **PAVS-VRF-05**(모델 진본성 — 선언 대조), **PAVS-VDT-07**(소급 감사)

**사건 시기**: 2026-09

**문제 상황.** 로봇 팔 집기 시험에서 그리퍼 몸체(손바닥)가 물체를 누르는 접촉은 충돌로 계수하고, 손가락이 물체에 닿는 접촉은 의도된 집기로 면제하였다. 다섯 기종을 같은 과제로 시험한 결과, 한 기종만 25 회 실행 전부에서 충돌 0 건을 기록하여 다섯 기종 중 유일하게 판정 계통의 충돌 문항을 통과하였다.

**무엇이 조용히 틀렸나.** 접촉의 역할(손바닥·손가락·팔)을 가르는 분류기가 코드에 내장된 손바닥 부위 이름 목록만 참조하였다. 그 기종의 그리퍼 몸체 부위 이름은 목록에 없었고, 부위 경로에 그리퍼 제품명이 들어 있다는 이유로 손가락으로 분류되었다. 손가락과 물체의 접촉은 의도된 집기로 면제되므로 25 회 전부가 계수되지 않았다. 실측 접촉력은 다른 기종의 손바닥 접촉과 같은 크기였고, 손가락 집는 힘의 약 10 배였다. 그 기종의 로봇 선언은 2026-09-06 부터 해당 부위를 몸체로 명시하고 있었고, 같은 체계의 다른 계측(하강 목표 제한)은 그 선언을 읽고 있었다. 한

사실을 두 곳이 따로 적었고, 분류기 쪽만 선언을 보지 않았다. '충돌 0 건'은 안전의 결과가 아니라 분류 규칙의 결과였다.

**무엇이 잡았나.** '한 기종만 충돌 0 건'이라는 미해결 질문을 안전의 증거로 읽지 않고, 접촉력·부위의 계층 위치·로봇 선언 세 축으로 대조한 절차. 세 축이 모두 '그 부위는 손바닥'을 가리켰다. 분류기가 로봇 선언을 읽도록 고치되, 선언된 이름은 그리퍼 하위 구조 안에서만 인정하였다 — 같은 이름이 팔 베이스 부위에도 흔히 쓰여, 이름만 보면 팔 베이스가 손바닥으로 분류되기 때문이다. 반례 시험(팔 베이스는 선언이 있어도 팔로 분류 · 선언이 없는 기종은 기존 목록으로 무회귀)으로 고정한 뒤 기준선을 다시 동결하였다. 재동결에서 해당 기종의 접촉은 25 회 전부 계수되었고, 집기를 충돌로 세는 오탐은 전 기종 0 건, 엄격 성공 지표는 전 기종 불변이었다. 재동결 직후, 판정 일부를 오프라인 재생으로 계수하는 도구가 같은 선언을 전달받지 못하고 있었음이 드러났다. 도구는 이튿날 고쳤으나 그 도구가 이미 낸 동결 판정의 재판정이 누락되었고, 그다음 날 다른 작업 중의 대조로 발견되어 해당 판정을 통과에서 한 단계 낮은 등급으로 정정하였다.

**일반화.** 판정 규칙이 '코드에 적힌 표'를 참조하면, 표에 없는 실물은 오류 없이 기본 분류로 떨어지고, 그 기본 분류가 면제이면 거짓 합격이 된다. 분류의 근거는 코드 내장 목록이 아니라 피시험체의 선언이어야 하며, 선언을 읽을 때는 그 선언이 유효한 범위(이 사례에서는 그리퍼 하위 구조)를 함께 제한하여야 한다. 한 사실을 여러 소비처가 쓰는 경우, 정본을 옮긴 뒤에는 소비처 전부가 따라왔는지 확인하고, 따라오지 않은 소비처가 낸 기존 판정은 소급하여 재판정한다. 표 밖 실물을 넣은 반례 시험이 없으면 이 결함은 '0 건'이라는 가장 좋은 성적으로 숨는다.

## 사례 C7. 값은 있는데 측정 증명이 없었다

**관련 조항**: **PAVS-VDT-01**(3 상 판정 — 필수 문항의 계측 부재 처리), **PAVS-MEA-01**(출처 증명 · 비실측 채점 금지)

**사건 시기**: 2026-09

**문제 상황.** 소형 이족 보행 로봇 3 대가 실내를 순찰하는 시험에서, 시험 담당자가 문항 7 건의 값을 기준값과 손으로 대조하여 '7/7 통과'로 보고하였다. 정식 판정 계통은 같은 실행을 9 문항 중 4 건 합격·5 건 불합격으로 판정하였다.

**무엇이 조용히 틀렸나.** 불합격 5 건 중 1 건은 실제 충돌이었고, 4 건의 사유는 같았다 — '이 지표에는 값이 있으나 이 실행에서 측정되었다는 기록이 없다. 값의 존재는 측정의 증거가 아니며, 필수 문항이므로 불합격'. 원인은 계측 결과를 결과 파일에 합치는 경로 세 곳(로봇별·로봇 간 관계 지표, 접촉 감시 지표, 센서 기록 계수)이 측정 증명 기록에 남기지 않고 값만 실은 데 있었다. 계측은 실제로 돌았으나 그 사실이 증명되지 않았다. 여러 대의 로봇과 접촉 감시 지표에 문항을 건 첫 시험이어서 그 전까지 드러나지 않았다. 사람의 대조는 값이 있다는 사실을 측정되었다는 사실로 읽었다.

**무엇이 잡았나.** 값이 아니라 측정 증명을 요구하고, 불합격에 사유를 남긴 정식 판정 계통. Part 1 v1.2 의 **PAVS-VDT-01** 이 허용한 처리 — 필수 문항의 계측 부재를 '계측 부재' 사유의 불합격으로 계상 — 가 그대로 작동하였다. 사유가 식별되어 있었으므로 이 불합격은 피시험체의 실패가 아니라 계측 배선의 결함으로 귀속될 수 있었다. 세 경로 모두 계측이 실행되는 지점에서 측정을 선언하도록 고쳤고, 가드를 신설하였다 — 문항이 '값은 있는데 측정 증명 기록에 없는' 지표를 쓰면 그 실행을 실패시키되, 계측기가 돌지 않은 경우(값 자체가 없음)와 배선 결함(값은 있고 증명이 없음)을 구별한다. 첫 경로를 고친 뒤 이 가드가 나머지 두 경로를 찾아냈다. 수정 전후 같은 시나리오의 증명 원장을 대조하여 해당 지표가 '측정'으로 기록됨을 확인하였고, 예제 전수 회귀를 통과하였다.

**일반화.** 판정의 입력은 값의 존재가 아니라 측정의 증명이다. 값과 기준값을 비교하는 것만으로는 판정이 성립하지 않으며, 측정 증명을 보지 않는 사람의 수동 대조는 관대한 방향으로 틀린다. 증명이 없는 값의 불합격에는 사유가 붙어야 한다 — 사유가 없으면 피시험체의 실패와 계측 배선의 결함이 같은 '불합격'으로 보이고, 사유가 있으면 고칠 곳

이 지목된다. 미측정을 합격으로 접지 않는 규칙(사례 B2)과 불합격의 사유를 식별하는 규칙은 함께 있어야 쓸모가 있다.

## D. 검증 — 채점기를 채점한다

### 사례 D1. 채점기 자신이 시험을 본다

**관련 조항**: **PAVS-VRF-01**(음성 대조군)

**사건 시기**: 2026-07

**문제 상황.** 교전 규칙 준수를 채점하는 문항군이 안정적으로 운영되고 있었다 — 다만 채점기의 위반 검출 능력은 별개의 검증 대상이다.

**무엇이 조용히 틀릴 수 있었나.** 이 사례는 사고가 아니라 예방 장치의 기록이다 — 검증하지 않으면, 채점기가 관대해지는 회귀는 합격만으로 구성된 스위트를 탐지 없이 통과한다.

**무엇이 잡았나.** 일부러 위반하는 시나리오(승인 없는 교전)를 회귀 스위트에 상주시켰다. 이 시나리오가 불합격으로 — 판정 근거에 해당 위반 문항이 남은 채로 — 판정되어야 스위트 전체가 통과가 되도록 하였고, 2026-08 기준 기대 판정과 실제 판정의 자동 대조가 스위트를 실행할 때마다 수행되었다. 다만 그 스위트를 주기적으로 자동 실행하는 연결은 미비했다(**PAVS-VRF-16** · 사례 D6).

**일반화.** 채점기의 변별력은 주장이 아니라 상설 시험의 대상이다. 합격만 있는 회귀 스위트는 채점기의 절반만 검증한다.

### 사례 D2. 재측정 규율이 규율 제정자 본인의 회귀를 검출했다

**관련 조항**: **PAVS-VRF-02**(골든셋), **PAVS-VRF-11**(변경 영향)

**사건 시기**: 2026-08

**문제 상황.** 측정 코드에 사소해 보이는 정렬 순서 변경이 들어갔다. "순서는 판정을 바꾸지 않는다"는 판단과 함께.

**무엇이 조용히 틀렸나.** 그 판단은 조건의 형태만 보고 내린 단정이었다. 실제로는 측정 경로의 길이가 바뀌어, 한 기종의 정확도 측정값이 15 배 악화되었다.

**무엇이 잡았나.** "계측 코드가 바뀌면 전 기종을 재측정하고 스냅샷과 대조한다"는 규율. 재측정이 회귀를 즉시 드러냈고, 원복 후 비트 단위 재현으로 완전성을 확인했다. 교훈이 함께 등기되었다: **"무관하다"는 판단은 "무엇의 함수인가"를 코드로 확인한 뒤에만 내린다.**

**일반화.** 채점·계측 코드의 변경은 그 코드를 분모로 쓰는 기존 판정의 재평가를 동반해야 한다. 검증 체계는 자기 자신의 변경에 대해서도 같은 규율을 적용한다.

### 사례 D3. 반증 실험의 독립변수가, 두 군 모두에서 바뀐 적이 없었다

**관련 조항**: **PAVS-VRF-03**(실효값 확인)

**사건 시기**: 2026-07

**문제 상황.** 물리 갱신 주기의 효과를 검증하는 A/B 실험이 "차이 없음 — 가설 반증"으로 끝났고, 그 결론 위에 운영 규칙까지 세워졌다.

**무엇이 조용히 틀렸나.** 별개의 가드 결함이 두 시험군 모두의 실효 주기를 같은 값으로 강등시키고 있었다. 독립변수가 한 번도 변하지 않은 실험이었다 — 반증이 아니라 무효였다. 오염은 수 주간 지속되었다.

**무엇이 잡았나.** 비교 시험의 독립변수를 설정 의도가 아니라 로그의 실효값으로, 각 군에서 개별 확인하는 규칙.

**일반화.** 실효값이 확인되지 않은 비교 시험은 반증이 아니라 미검증이다. 잘못된 반증은 잘못된 규칙을 산출하며, 그 규칙은 장기간 유지된다.

## 사례 D4. 계측 버그가 로봇 결함으로 위장했다

**관련 조항**: **PAVS-VRF-04**(계측기 먼저 반증), **PAVS-OPS-02**(정정 보존)

**사건 시기**: 2026-07

**문제 상황.** 집기 실패가 반복되자 "그리퍼 자산의 기하 결함"으로 판정되어 자산 재제작 지시 직전까지 갔다.

**무엇이 조용히 틀렸나.** 실제 근본 원인은 두 개의 계측·매핑 버그였다 — 그리퍼 간격 변환식의 오적용으로 그리퍼가 과주행했고, 집는 힘 계측이 무관한 접촉 콜백에 0 을 덮어써 진짜 집기를 "잼"으로 기각하고 있었다.

**무엇이 잡았나.** 피검체 귀속 전에 계측 계통을 반증하는 절차 — 간격 변환의 독립 검산과 집는 힘 계측의 대조 시험 — 가 두 버그를 드러냈고, 계측 수정 후 같은 자산이 연속 합격했다. 원 판정은 데이터로 철회되었고, 철회 기록은 지워지지 않고 정정 연결로 보존되었다.

**일반화.** 시험 실패를 피시험체에 귀속하기 전에 계측 계통을 먼저 반증한다. 계측 결함은 피시험체 결함으로 위장하는 성질이 있다 — 그리고 오진의 이력 자체가 검증 자산이므로 지우지 않고 정정 연결로 보존한다.

## 사례 D5. 안전 가드가 한 번도 발화한 적이 없었다

**관련 조항**: **PAVS-VRF-13**(가드 발화 검증)

**사건 시기**: 2026-08

**문제 상황.** 자사 연구용 **격리 실험 셀**(작업자가 함께 있는 협동 환경이 아니다)에서 로봇 팔을 시험하고 있었다. 이 셀은 하드웨어 비상정지가 없어 소프트웨어가 안전의 본체이고, 접촉 감지 가드가 "켜져" 운영되고 있었다.

**무엇이 조용히 틀렸나.** 가드가 기준선을 읽는 경로가 토크 값을 채우지 않는 호출이어서, 가드는 매번 내부 오류로 중단되고 있었다 — 즉 접촉 정지는 운영 기간 내내 **한 번도 작동한 적이 없었다.** 켜져 있다는 사실과 동작한다는 사실은 다른 명제였다.

**무엇이 잡았나.** 수정 후, 기준값을 낮춰 실제 발화를 주입 시험으로 확인하고, 위약 대조(발화하면 안 되는 조건에서의 오발 0 건)까지 검증했다. 적대 검토가 추가 함정도 잡았다: 고정 기준선은 자세에 따라 중력을 접촉으로 오인할 수 있다 — 민감도는 상수가 아니라 자세의 함수였다.

**일반화.** 안전 장치의 검증은 "호출되는가"가 아니라 "호출되면 실제로 동작하는가"까지다: 주입 시험으로 발화를, 위약 대조로 오발과 둔감을 확인한다.

## 사례 D6. 검사기는 존재하였으나 실행되지 않았다

**관련 조항**: **PAVS-VRF-16**(자기 검사의 상시 실행), **PAVS-VRF-11**(변경 영향 분석)

**사건 시기**: 2026-08

**문제 상황.** 채점 체계에 자기 검사 장치가 갖추어져 있었다. 판정면에 빈 입력을 주입하여 합격 산출 여부를 확인하는 프로브, 낡은 경계 설명을 지목하는 탐지기, 문항의 변별력을 다시 계산하는 도구. 각각은 자기검증(셀프테스트)까지 통과했고, 결함을 주입하면 실제로 발화했다.

**무엇이 조용히 틀렸나.** 그 도구들이 **어떤 자동 실행 지점에도 연결되어 있지 않았다.** 주기 실행 없음, 통합 검사에서의 실행 주체 없음. 유일한 자동 실행 지점은 변경 시점의 훅 하나였는데, 그 훅은 경로 범위로 게이팅되어 있어 프로브가 보호하는 판정면의 대부분을 검사 대상에서 제외하고 있었다. 즉 **가드의 자동 커버리지가 그 가드가 지키려는 위험면과 역상관**이었다. 이 상태에서 산출되는 '위반 0'은 적합의 결과가 아니라 미실행의 결과였으며, 두 상태가 보고서에서 동일하게 표시되었다.

**무엇이 잡았나.** 규격 적합성에 대한 자체 점검. 구현 여부가 아니라 **실제 실행 여부**를 기준으로 삼자 존재와 작동이 구별되었다. 이 기준은 도구를 개발한 주체가 스스로 적용하기 어렵다. 개발 주체에게는 장치의 존재가 곧 작동으로 인식되기 때문이다.

**일반화.** 검사 장치의 존재는 그 검사가 수행되었다는 증거가 아니다. 자기 검사는 주기 또는 변경 시점에 자동으로 실행되어야 하고, 실행 범위가 보호 대상 전체를 덮지 못하면 제외된 범위를 밝혀야 한다. 실행되지 않은 검사의 '위반 없음'은 적합의 결과가 아니라 미실행의 결과다.

## 사례 D7. 제조자가 같은 항목에 두 값을 내고 있었다

**관련 조항**: **PAVS-VRF-05**(모델 진본성)

**사건 시기**: 2026-08

**문제 상황.** 시뮬레이션 모델의 제원이 제조자 사양과 맞는지 자동으로 대조하는 도구를 만들고 있었다. 한 로봇 팔에서 관절 한계가 제조자 공표값보다 6/7 관절에서 좁게 들어가 있는 것이 잡혔고, 최대 차이는 9 도였다. 좁은 쪽이라 거짓 합격을 만들지는 않으나 능력의 과소 보고가 되므로 결함 후보로 기록했다.

**무엇이 조용히 틀렸나.** 세 번째 기종을 같은 방법으로 대조하다가 전제가 무너졌다. 그 제조자의 공식 배포 자료가 한 관절의 한계를 **하드웨어 한계의 절반**으로 적으면서, 그 사실과 사유를 자료 안에 함께 적어 두고 있었다. 인접 관절이 간섭해 물리적으로 그 이상 회전할 수 없고, 운동 계획이 불연속 해공간에서 실패하는 것을 피하기 위한 축소였다.

**즉 '제조자 확정값'이 하나가 아니었다.** 하드웨어 한계와, 계획을 위해 축소한 기술서 값이 둘 다 같은 제조자에게서 나온다. 어느 쪽을 기준선으로 삼느냐에 따라 같은 모델이 「180 도 좁은 결함」이 되기도 하고 「정확히 일치」가 되기도 한다.

값의 일치만 보는 대조는 이 두 기종을 **정반대로 틀리게** 판정한다 — 근거가 기록된 축소는 결함으로, 근거 없는 축소는 「조금 좁을 뿐」으로.

**무엇이 잡았나.** 대조 대상을 늘린 것이 잡았다. 두 기종까지는 「우리 값이 좁다」가 일관된 결함으로 보였고, 세 번째에서야 그 해석이 기준선의 선택에 의존한다는 것이 드러났다. 두 기종의 차이는 축소의 크기가 아니라 **근거의 유무**였다 — 한쪽은 제조자가 사유를 적어 두었고, 다른 쪽은 어느 자료에도 그 축소를 만드는 표현이 없었다.

**일반화.** 모델 진본성 검사가 보아야 하는 것은 값의 일치가 아니라 **차이의 설명 가능성**이다. 제조자가 같은 항목에 복수의 값을 낼 때는 채택한 기준선과 그 사유를 선언하고, 모델이 기준선과 다른 항목마다 근거를 기록한다. **근거가 기록되지 않은 차이만이 미검증이다.** 차이의 방향(넓음/좁음)은 위험도가 다르므로 함께 남긴다.

## 사례 D8. 자산이 선언한 관절 한계를 로더의 포괄 규칙이 428 배로 덮었다

**관련 조항**: **PAVS-VRF-05**(모델 진본성 대조)

**사건 시기**: 2026-07

**문제 상황.** 휴머노이드 로봇이 시험 중 반복적으로 넘어졌다. 원인을 찾는 과정에서 손 부위 14 개 관절의 구동 한계를 확인하였다.

**무엇이 조용히 틀렸나.** 자산 파일이 저작한 값은 0.7 N·m 였으나, 시뮬레이션에 실제로 적재된 값은 300 N·m 였다 — **428 배**다. 자산의 선언을 로봇 등록부의 포괄 규칙(관절 이름이 특정 문자열로 끝나면 300 을 적용)이 덮어쓴 결과였다. 오류는 발생하지 않았고 로봇은 정상적으로 적재되어 시험을 수행하였다. 사례 D7 이 **제조자가 두 값을 낸 경우**를 다룬 데 비하여, 이 사례는 **자산의 값이 하나뿐인데도 우리 적재 계통이 다른 값을 넣은 경우**다.

**무엇이 잡았나.** 시뮬레이션에 적재된 값과 자산이 선언한 값의 대조. 이 대조가 없으면 시험은 '선언한 로봇'이 아니라 '적재된 로봇'에 대하여 수행되며, 성과 문서는 전자의 이름으로 작성된다.

**일반화.** 모델 진본성의 위협은 외부 출처의 불일치뿐 아니라 **자기 적재 계통의 조용한 덮어쓰기**에서도 온다. 기본값을 적용하는 포괄 규칙은 명시적 선언을 이길 수 없어야 하며, 적재된 값과 선언된 값의 대조는 시험 개시 조건이어야 한다. 다만 이 사례에서 해당 수정은 **증상을 바꾸지 못하였다** — 수정 전후로 관절의 물리 응답이 소수점까지 동일하였다. 결함의 수정은 그 자체로 원인의 규명이 아니며, 두 차례의 수정이 증상을 바꾸지 못한 시점에서 가설이 아니라 접근을 바꾸어야 한다.

## 사례 D9. 한 기종을 고친 수정이 이미 작동하던 두 기종을 무너뜨렸다

**관련 조항**: **PAVS-VRF-11**(변경 영향 분석 · 기존 판정 재평가)

**사건 시기**: 2026-07

**문제 상황.** 휴머노이드의 지형 감지 입력에서 센서 원점 보정값이 누락되었음이 확인되어, 이를 보정하는 수정이 적용되었다. 수정 자체는 해당 기종에서 검증되었다.

**무엇이 조용히 틀렸나.** 같은 감지 계통을 공유하는 다른 다리 로봇 2 종(사족 1 종·이족 1 종)이 그 수정으로 **보행이 무너졌다**. 두 기종은 수정 이전까지 정상 작동하고 있었다. 수정은 대상 기종만을 보고 검증되었고, 같은 계통을 소비하는 다른 기종의 판정은 재평가되지 않았다.

**무엇이 잡았나.** 수정은 전면 원복되었고, 원복 사실과 그 사유가 해당 계통의 문서에 '이 수정을 다시 시도할 경우 함께 읽어야 하는 기록'으로 지정되었다.

**일반화.** 공유 계통에 대한 수정은 그 계통을 소비하는 **전 기종의 판정을 함께 이동시킨다**. 한 기종에서의 검증은 변경의 검증이 아니라 변경의 표본이다. 로봇 비종속을 주장하는 채점 체계에서 이 성질은 강점인 동시에 최대의 회귀 경로이며, 공유 계통의 변경에는 소비 기종 전수의 재평가가 선행하여야 한다.

## 사례 D10. 계측용 '뷰'가 피시험체의 물리를 바꾸었다

**관련 조항**: **PAVS-VRF-02**(골든셋 판정 보존), **PAVS-VRF-11**(변경 영향 분석), **PAVS-MEA-06**(신설 채널 실산출 확인)

**사건 시기**: 2026-09

**문제 상황.** 그리퍼 몸체(손바닥)가 물체를 누르는 접촉을 재려고, 손바닥 부위에 물리 상태를 읽는 계측용 '뷰'를 추가하였다. 다음 날 전면 재측정에서 한 기종의 삽입 시험 칸(기종과 과제의 조합 하나로 이루어진 시험 단위)만 기준선과 달랐다 — 삽입 위치로 하강하는 동안 팔 끝이 좌우 ±4 cm 로 주기 진동한 뒤 삽입은 성공하였다. 엄격 성공 지표만 보면 아무것도 달라지지 않았다.

**무엇이 조용히 틀렸나.** 뷰를 만드는 편의 기능은 대상 부위에 강체 물리 속성이 없으면 그 속성을 부여하는 동작을 함께 수행하였다. 그 기종의 손 부위는 모델을 가져올 때 고정 관절로 부모 부위에 병합되어 강체 속성이 없었고, 뷰가 만들어지는 순간 별개의 강체로 바뀌었다. 계측 장치가 피시험체의 물리를 바꾼 것이다(사례 B4 '채점기가 피시험체를 느리게 만들었다'와 같은 계열). 이미 강체 속성을 가진 부위에 뷰를 만든 다른 기종의 결과는 바이트 단위로 동일하였다 — 원인은 뷰의 생성이 아니라 없던 물리 속성의 부여였다.

**무엇이 잡았나.** 물체 자세가 고정된 시험 칸의 재측정은 기준선의 복제여야 한다는 규율(사례 D11)이 감시 도구에 있었다. 복제가 깨진 시험 칸 하나를 뷰만 끄고 한 번 더 실행하자 진동이 사라졌고(궤적 반전 2 건/1.38 m, 기준선과 동일), 뷰를 자세만 읽는 방식으로 바꾼 확인 실행은 뷰를 끈 실행과 바이트 단위로 같았다. 같은 결함의 재발을 막는 자기 시험을 두었다.

**일반화.** 계측은 피시험체의 물리를 바꾸어서는 안 된다 — 읽기 전용 접근을 기본으로 하고, 속성이 '없으면 부여하는' 편의 기능은 피시험체 위에서 쓰지 않는다. 계측을 추가한 뒤에는 **같은 입력의 재측정이 이전 결과의 복제인지**(복제가 성립하지 않는 시험 칸이면 알려진 변동 범위 안인지 · 사례 D14) 확인하는 것이 계측 변경의 완료 조건이다. 신설 계측의 실산출 확인(**PAVS-MEA-06**)과 변경 뒤의 골든셋 대조(**PAVS-VRF-02**)·변경 영향 분석(**PAVS-VRF-11**)이 이 확인을 일부 덮으나, '계측이 피시험체를 바꾸지 않아야 한다'는 요구 자체를 명시한 조항은 없다 — 사례 C5 와 같이 조항 확장 후보로 Part 1 12.6 의 차기 정식 개정 검토 대상에 포함하였다.

## 사례 D11. 자세 고정 시험의 반복 실행은 표본이 아니라 복제였다

**관련 조항**: **PAVS-REP-06**(통계 규율 — 시행수와 불확도), **PAVS-ITM-09**(섭동 축 변주), **PAVS-VRF-14**(재현 · 구성 고정), **PAVS-VDT-01**(3 상 판정 — '못 가름' 상태 유지)

**사건 시기**: 2026-09

**문제 상황.** 단일 실행으로는 궤적 반전 사건이 적어 신뢰구간이 기준값에 걸쳐 '못 가름'(합격·불합격 어느 쪽으로도 판정할 수 없는 상태)으로 남는 시험 칸(기종과 과제의 조합 하나로 이루어진 시험 단위)이 있었다. 표본을 늘리려고 같은 시험 칸을 네 번 더 실행하여 합산하려 하였다.

**무엇이 조용히 틀렸나.** 그 시험 칸은 물체 자세가 고정되고 물리 계산이 결정적이어서, 반복 실행은 서로 다른 계산 노드에서 수행하여도 궤적 기록이 **바이트 단위로 같았다**(접촉이 있는 시험 칸도 39 μm 이내). 같은 궤적을 다섯 번 세어 사건 수 기반의 신뢰구간을 좁혔다면 '표본이 늘었다'는 서술은 거짓이 되었을 것이다.

**무엇이 잡았나.** 합산 도구에 '궤적의 점별 편차가 0.1 mm 이내이면 복제로 보고 합산에서 제외'하는 규칙을 두었고, 반복 실행 열한 건 전부가 제외되어 '못 가름' 판정은 그대로 유지되었다. 정보를 늘리는 길은 반복이 아니라 **한 실행 안에서 더 많은 운동을 측정하는 것**(정확도 시험 구간까지 궤적 기록에 포함)이었고, 그렇게 하자 같은 시험 칸이 단일 실행으로 판정되었다(반전 13 건/6.27 m, 신뢰구간이 기준값 아래).

**일반화.** 반복 횟수는 실행 간 독립성이 있을 때만 표본이다. 재현이 완벽한 시뮬레이션에서 '표본 수를 늘린다'는 말은 실행마다 무엇이 달라지는지(난수 시드·물체 자세의 변주 폭)를 먼저 밝혀야 성립한다. 복제는 표본이 아니라 **재현성 증거**로만 기록한다. 판정할 수 없는 상태를 복제로 메워 판정 가능한 상태로 만들지 않는 것은, 미측정을 합격·불합격으로 접지 않는 3 상 판정의 취지와 같다.

**후속(2026-10).** 이 사례의 시험 칸은 실제로 바이트 단위로 같았다. 그 뒤 같은 조건(같은 코드 · 같은 구성)의 고정 자세 재실행을 다시 모아 보니 궤적이 점 단위로 같은 묶음은 39 개 가운데 8 개였고, 한 시험 칸에서는 판정까지 갈렸다(사례 D14). 이 사례의 일반화는 '복제인 경우'에 대한 것이며, 복제인지는 가정하지 않고 실행마다 대조해 확인한다.

## 사례 D12. 팔 자체 무게 없이 정밀도를 쟀다

**관련 조항**: **PAVS-VRF-05**(모델 진본성 대조), **PAVS-MEA-07**(요청값·실효값·적용 병기), **PAVS-VRF-11**(변경 영향 분석 · 기존 판정 재평가), **PAVS-VRF-02**(골든셋 판정 보존)

**사건 시기**: 2026-10

**문제 상황.** 한 산업용 로봇 팔의 시뮬레이션 모델이 위치 정확도·집기·배치·삽입의 네 시험에서 측정되어 기준선으로 동결되어 있었다. 이 모델은 공개 배포된 원본 모델에 그리퍼를 결합하여 저작한 것이었다. 모든 실행이 오류 없이 정상 종료되었다.

**무엇이 조용히 틀렸나.** 원본 모델의 팔 링크 9 개 전부가 중력 비활성으로 저작되어 있었고, 결합 모델이 그 설정을 그대로 물려받았다. 그리퍼와 물체에는 중력이 작용하였으나 팔 자체에는 무게가 없었다. 위치 정확도에는 팔 자체 중량에 의한 처짐이 들어 있지 않았다. 질량·관절 한계 같은 제원 값은 정상으로 보였고, 링크별로 중력이 실제로 적용되는지를 기록하거나 검사하는 단계는 체계 어디에도 없었다.

**무엇이 잡았나.** 정격 하중 시험 조건(사례 B7)을 추가하자, 하중을 부착할 링크의 중력이 꺼져 있어 하중을 더하여도 처짐이 생길 수 없었고, 시험은 조용히 근사하지 않고 실행을 거부하였다(fail-loud). 그 거부를 원인까지 추적하여 모델을 판독하자 팔 링크 전체의 결함이 드러났다. 조치는 네 가지였다 — 결합 모델의 중력 설정 정규화(저작된 속성·관계 2,241 개 중 변경은 중력 비활성 표지 9 개뿐임을 대조로 확인), 저작 단계에 정규화를 넣어 재저작 시의 상속 차단, 실행 시작 시 링크별 중력 상태를 기록하고 실행 제출 전에 모델 파일을 판독하는 계기 신설, 해당 네 시험의 재측정. 재측정 판정은 넷 모두 유지되었으나 수치는 이동하였다. 같은 계기로 다른 공개 모델을 판독하자 같은 결함이 다른 모델에서도 확인되어, 해당 모델의 소관 쪽에 통보하였다.

**일반화.** 모델 진본성 검사는 제원 값의 대조로 끝나지 않는다. 그 제원에 물리가 실제로 적용되는지 — 중력·충돌체·구동 한계가 시뮬레이션 안에서 작동하는지 — 까지 포함하여야 한다(사례 C4·D7·D8 이 값의 불일치를 다룬 데 비하여, 이 사례는 값은 맞고 물리가 꺼져 있던 경우다). 재측정에서 판정이 유지되었다는 사실은 결함이 이번에는 판정을 바꾸지 않았다는 사후 확인일 뿐, 결함 기간의 측정을 유효하게 만들지 않는다. 정규화의 변경 범위를 대조로 한정하여 다른 속성이 바뀌지 않았음을 보인 것은 변경 영향 분석의 최소 단위다. 이 결함은 새 시험 조건이 처음으로 그 물리량에 의존한 순간에 드러났다 — 새 조건이 낸 거부는 우회할 대상이 아니라 모델의 숨은 결함을 가리키는 신호로 추적하여야 한다.

## 사례 D13. 사전등록 게이트는 실행이 끝난 뒤에 불렸다

**관련 조항**: **PAVS-VRF-16**(자기 검사의 상시 실행), **PAVS-OPS-06**(게이트=실측 근거 · 거부 가시 통지), **PAVS-VRF-06**(정합=측정 대상 · 보류 시험 채택), **PAVS-OPS-02**(등기 수명주기 · 정정 보존)

**사건 시기**: 2026-09

**문제 상황.** 비교 실험은 실행 전에 가설·판별자·난수 시드·실행 순서를 사전등록 파일로 고정하고, 검사 게이트가 그 파일의 형식 규칙을 확인하도록 되어 있었다. 시험 대상 표기 형식·명시 시드 목록·실행 순서 선언의 세 규칙은 2026-09-06 에 도입되었다.

**무엇이 조용히 틀렸나.** 2026-09-16 에 등록된 사전등록 파일 하나가 세 규칙을 모두 위반한 채로 실행되어, 이틀 동안 340 회의 시험 실행을 소비하였다. 그 파일은 게이트를 통과한 적이 없었다. 게이트는 배치가 끝난 뒤 가설 판정표를 작성하는 단계에서만 호출되었고, 실행 제출 전의 검증은 없었다. 처음의 진단은 이 1 건을 '규칙 이전 기록'으로 보고 유예(소급 적용 면제) 여부를 결정하려 하였으나, 이력을 대조하자 규칙이 파일보다 10 일 앞서 있었다. 유예의 문제가 아니라 미집행의 문제였다. 같은 시기 작업 공간에는 결과가 나온 뒤 그 파일의 판별자를 바꾸는 미확정 편집이 남아 있었다.

**무엇이 잡았나.** 예외 1 건을 유예로 처리하기 전에 규칙 도입 시점과 파일 등록 시점을 이력으로 대조한 절차. 게이트를 실행 제출 앞으로 옮겨, 위반이면 실행 0 건 상태로 배치를 거부하게 하였다. 알고도 실행하여야 하는 경우에는 8 자 이상의 사유를 남겨야 하며, 그 사유는 실행 구성에 함께 기록된다 — 조용한 우회는 허용하지 않는다. 게이트 호출이 가설 판정표 작성보다 앞선다는 순서 자체를 시험으로 고정하였다. 위반 파일은 확정된 원본으로 복원하고 고치지 않았다 — 옛 기록을 새 규칙에 맞추어 고치면 그때 무엇이 등록되어 있었는지가 사라진다.

**일반화.** 검사가 배선되어 있다는 것과 검사가 집행된다는 것은 다르다. 사례 D6 이 '어디에도 연결되지 않은 검사'였다면, 이 사례는 **연결은 되어 있으나 위치가 틀린 검사**다 — 실행 이후에 불리는 사전 검사는 위반을 막지 못하고 사후에 기록할 뿐이다. 검사의 유효성은 존재 여부가 아니라 호출 시점 — 보호하려는 실행보다 먼저 불리는지 — 으로 판별하며, 그 호출 순서는 시험으로 고정하여야 한다. 규칙 이전 기록의 유예와 규칙 이후의 미집행은 겉으로 같아 보이므로 시점 대조로 가른다. 결과를 본 뒤 사전등록을 고치는 것은 사전등록의 무효화이며, 옛 기록은 위반 상태 그대로 보존하고 정정은 별도 기록으로 연결한다.

## 사례 D14. 같은 자세의 재실행이 판정을 바꾸었다

**관련 조항**: **PAVS-VRF-14**(재현 · 구성 고정), **PAVS-VRF-09**(결정적 재현은 시스템 결함으로 우선 취급), **PAVS-REP-06**(통계 규율 — 시행수와 불확도)

**사건 시기**: 2026-10

**문제 상황.** 물체 자세가 고정된 끼워 넣기 시험 칸 하나에서, 같은 코드 · 같은 구성 · 같은 자세로 실행한 결과가 두 경로로 나뉘었다. 한 경로에서는 부품이 구멍 입구에서 그대로 미끄러져 들어갔고, 다른 경로에서는 입구에 끼어 걸림 탐색을 아홉 번 한 뒤에 들어갔다. 두 경로 모두 삽입에는 성공하였으나, 탐색 왕복이 궤적 반전으로 계상되어 시험 칸 판정은 '합격'과 '주의'로 갈렸다. 기준선은 앞서 측정한 기준 자세 실행 한 건을 다시 쓰고 있었고, 그 한 건은 좋은 경로였다.

**무엇이 조용히 틀렸나.** 이 시험 칸의 반복 실행은 복제라고 가정되어 있었다(사례 D11). 그 가정 아래에서는 기준 자세를 회차마다 다시 잴 이유가 없었고, 다시 쓴 실행 한 건은 사실상 두 경로 가운데 하나를 뽑은 결과로 기준선에 고정되었다. 같은 물리 궤적을 내는 코드 판들의 기준 자세 실행 11 건을 모으자 나쁜 경로가 3 건이었다. 나쁜 경로끼리는 서로 다른 코드 판 · 다른 날의 실행에서도 궤적이 점 단위로 같았다 — 무작위 잡음이 아니라 결정적인 경로 두 개였다.

**무엇이 잡았나.** 갈림을 환경이나 동시 실행 탓으로 돌리지 않고(**PAVS-VRF-09**) 단독 순차 실행으로 갈림 지점을 좁혔다. 관절과 물체 상태를 비트 단위로 남기는 계측을 넣자, 두 경로는 손에 쥔 부품을 카메라로 관측하기 직전까지 상태가 같았고 그 관측에 쓴 깊이 영상의 지문만 달랐다. 깊이 영상이 실행(프로세스)마다 마지막 자리에서 다르게 계산되어 부품 끝의 관측값이 22 nm 달라졌고, 정렬과 첫 접촉이 이 차이를 미끄러짐과 끼임의 분기로 바꾸었다. 인과는 관측값만 바꿔 끼우는 대조로 확정하였다 — 좋은 관측을 낸 실행에 나쁜 경로의 관측 비트만 덮자 나쁜 경로가 궤적까지 그대로 재현되었다. 같은 실행 안에서 같은 장면을 열 번 다시 찍으면 깊이 영상은 매번 같았으므로, 갈림은 프레임이 아니라 실행 단위로 정해졌다. 저장된 결과물로부터 판정을 다시 계산하는 재현(**PAVS-VRF-14**)은 이 갈림과 무관하게 성립하였다 — 갈린 것은 채점이 아니라 시뮬레이션의 재실행이었다. 처치로는 관측값을 반올림해 결과를 한쪽으로 고정하는 방법은 쓰지 않았다. 실물이 같은 민감도를 가진다면 성적이 실물보다 좋게 나오기 때문이다. 시험 칸의 거동은 그대로 두고, 기준 자세 실행의 재사용을 폐지하여 회차마다 새로 재게 하였으며, 결과 문서에 '같은 자세로 다시 시험해도 결과가 달라질 수 있음'을 적고 시도별 걸림 탐색 횟수를 함께 표기하였다.

**일반화.** 사례 D11 의 교훈 — 복제는 표본이 아니다 — 은 복제일 때 성립한다. 시뮬레이션이라도 시드 · 구성 · 코드를 고정한 재실행이 같은 결과를 낸다는 것은 가정이 아니라 측정할 사실이다. 같은 조건의 고정 자세 재실행을 다시 모으자 궤적이 점 단위로 같은 묶음은 39 개 가운데 8 개였다. 기준선에 실행 한 건을 다시 쓰는 관행은 그 한 건이 추

첨의 결과일 수 있다는 사실을 숨긴다. 결과가 갈리는 시험 칸의 민감도는 없앨 대상이 아니라 재고 공시할 대상이며 (**PAVS-REP-06**), 갈림의 비율이 곧 그 시험 칸의 정보다.

## E. 보고 — 결과의 기술 방식

### 사례 E1. 같은 시험에 성적서 두 장 — 결과는 정반대

**관련 조항**: **PAVS-REP-02**(귀속·유령 표준), **PAVS-VDT-02**(긍정 증거), **PAVS-VDT-06**(권위 계층), **PAVS-REP-03**(커버리지 목록), **PAVS-REP-04**(자기선언 순환 금지)

**사건 시기**: 2026-08

**문제 상황.** 같은 시험 실행에 대해 두 경로가 성적서를 냈다: 하나는 불합격, 다른 하나는 "합격 — 운용 인증됨"(당시 표기).

**무엇이 조용히 틀렸나.** 후자의 합격은 시험 데이터가 만든 것이 아니었다 — **빈 입력으로 생성해도 완전히 동일한 성적서가 나왔다.** 부재를 0 으로 치환하고, 기본값을 생성하고, 요청자가 파라미터로 표준 이름을 지정하였다. 같은 감사에서 별건으로, 또 다른 발급 경로는 발행 체계 어디에도 등록되지 않은 표준 이름을 전 성적서에 기재하고 있었다 — 그 이름의 표준이 요구하는 측정은 수행된 적이 없었다.

**무엇이 잡았나.** 유령 표준 게이트(미등록 표준 이름의 발행 거부 — 실제 발행 시도에서 발화 확인)와 귀속 공개. 수정 뒤의 성적서는 문항 단위로 "권위 유래 1 건, 자체 정의 12 건"을 스스로 표기하였고, 종전 양식에서 '표준 통과'로 표시되었을 문서가 자기 한계를 명시하였다. (성적서 발급 기능은 이 판 기준 꺼져 있다.)

**일반화.** 성과 보고의 표준이란 곧 **양식이 스스로 자기 증거의 급을 말하게 하는 것**이다. 표준 이름은 요청자가 고르는 라벨이 아니라 등록·검증된 귀속이다.

### 사례 E2. 채점은 정확했으나 성과 문서가 반대 결과를 기술했다

**관련 조항**: **PAVS-REP-13**(판정 정본 직결)

**사건 시기**: 2026-08

**문제 상황.** 채점 체계는 해당 트랙의 통과를 기록했는데, 자동 생성된 성과 문서(PDF)는 같은 시험을 무관한 기준의 치명 불합격으로 서술했다.

**무엇이 조용히 틀렸나.** 문서 생성기가 판정 결과 파일을 열지 않고 있었다 — 파일 이름 불일치의 연쇄로, 빈자리를 무관한 도메인의 기준값 표가 채웠다. 채점이 정확하더라도 표시 계층이 병렬의 판정을 생성하면 최종 독자는 잘못된 결과를 읽는다.

**무엇이 잡았나.** 성과 문서를 판정 정본에서 직접 생성하도록 재배선하고(중간 재해석 금지), 문서에 '측정 신뢰도' 절 — 측정 항목, 미측정 항목, 유보가 필요한 값 — 을 자체 수록하도록 하였다.

**일반화.** 판정과 보고 사이에 재해석 계층을 두지 않는다. 보고서는 판정 정본의 렌더링이어야 하며, 자기 신뢰도를 스스로 공개해야 한다.

### 사례 E3. 데이터 절반을 삭제해도 '수집율 100%'가 산출되었다

**관련 조항**: **PAVS-REP-06**(통계·지표 검증), **PAVS-VRF-01**(대조군 원리)

**사건 시기**: 2026-07

**문제 상황.** 새로 구현된 데이터 수집 건전성 지표가 어떤 조건에서도 100%를 보고했다.

**무엇이 조용히 틀렸나.** 지표의 분모가 "수집된 것"이었다 — 절반을 지워도, 수집된 것 중 수집된 비율은 100%다. 지표가 자기 자신을 참조하면 어떤 손실도 검출되지 않는다.

**무엇이 잡았나.** 지표 자체에 대한 음성 대조: 알려진 결함(의도적 데이터 삭제)을 주입하고 지표가 그것을 검출하는지 자체 시험(selftest)으로 증명하게 했다 — 이 대조 시험이 출고 전 단계에서 거짓 지표를 검출하였다.

**일반화.** 성과 문서에 실리는 지표는 그 지표가 결함을 검출할 수 있음을 스스로 증명해야 한다. 분모가 자기 자신인 지표는 지표가 아니다.

### 사례 E4. 정확도 "개선"은 붕괴의 반전이었다

**관련 조항**: **PAVS-REP-06**(널대조군·붕괴 지표)

**사건 시기**: 2026-08

**문제 상황.** 판정 보조 모델의 미세조정이 정확도를 올렸다는 결과가 나왔다 — 균형 평가 세트(40 편 · 합격 20 · 불합격 20)에서 20/40(50.0% · Wilson 95% 35.2 ~ 64.8%) → 25/40(62.5% · Wilson 95% 47.0 ~ 75.8%) · 두 구간 겹침(2026-08).

**무엇이 조용히 틀렸나.** 조정 전 모델은 전부를 한 클래스로(예측 불합격 40 · 합격 0), 조정 후 모델은 거의 전부를 반대 클래스로(예측 합격 35 · 불합격 5) 답하고 있었다 — 평가 세트가 균형이라 **단일 클래스 붕괴의 방향만 뒤집혀도 산술적으로 그 개선폭이 나온다.** 손실 곡선은 정상처럼 보였다. 별건의 개선 주장에서는, 대조군 없이 주장되던 개선폭이 통계적으로 성립하지 않아 철회되었고, 동일 처리 경로만 통과시킨 널대조군을 세우자 비로소 개선분의 귀속이 유의하게 성립했다 — '처리 경로 자체의 효과'라는 반론도 그 대조군이 반박했다.

**무엇이 잡았나.** 정확도 단독 인용 금지 — 널대조군과 붕괴 지표(클래스 분포)의 병기 의무.

**일반화.** 개선 주장은 널대조군과 붕괴 지표를 병기해야만 성립한다. 손실이 내려갔다는 것과 무언가를 배웠다는 것은 다른 명제다.

## 맺음 (informative)

이 사례들에서 반복되는 구조는 하나다: **검증되지 않은 가정 — 값은 측정일 것이다, 문항은 재고 있을 것이다, 가드는 동작할 것이다, 성과 문서는 판정을 반영할 것이다 — 이 조용히 무너지는 것.** Part 1 의 각 조항은 그 가정 하나씩을 명시적 검증 의무로 바꾼 것이다.

사례의 증가는 이 규격이 유지·개정되고 있음을 의미한다. 새 사고는 새 조항의 후보이며, 개정판 사례집에 추가 수록한다.

*PAVS Part 2 사례집 발행본 v1.3 · 2026-10-29 · 수록 39 건(문항 7 · 계측 7 · 판정 7 · 검증 14 · 보고 4). 각 사례의 원 기록과 증거는 발행 주체가 보관하며, 공동 시범 시험 상대에게 제공한다.*