

결론 — 나는 틀린 진단을 데이터로 뒤집었다

확정 v1(2026-07-10) → v1.1(2026-07-26): 게이트 C(설문 파일럿) 반영 — 요약 1행, \$2 후속 검증 절, \$3 소구 축 실물, \$4 한계, \$5 규율 항목 추가. 기존 판정(가설 반박 3단 반전·방향 판정)은 무변경 — 게이트 C는 이를 보강했다.

한 장 요약

항목	내용
사건	완벽히 준비한 7분 본선 PT, 그러나 탈락 (심사위원 질문 0·공식 피드백 0)
세운 가설(H2)	“차별점으로 매실청(A)을 <i>과대</i> 강조한 것이 탈락 원인. 완성도·외식 대체(C)를 앞세웠어야 했다”
검증 방법	강조점만 다른 3-arm 가짜 상세페이지(fake-door) — T1(완성도)·T2(매실청)·C0(대조군), 판정 기준 사전 등록
결과	대조군 대비 전환율 T2(매실청) 81.5% > T1(완성도) 48.5% → 가설 반박
판정	매실청이 오히려 더 나은 과녁이었다. 문제는 매실청을 세운 것이 아니라, 그것을 충분히 세우지 않고 완성도를 앞세운 조준이었다 실제 탈락 원인은 확정 불가. 단, 서류를 통과시킨 강점(매실청)을 본선에서 3시나리오로 희석했을 가능성 — 원인 후보가 ‘무엇을 골랐나(방향)’에서 ‘어떻게 소구했나(실행)’로 이동. 교훈: 소구는 뽕족하게, 방어는 두껍게
함의(왜 떨어졌나)	설문 파일럿(사전등록, 유효 N=88)으로 남은 공백을 잰다 — 항 기피 수요층 33% [24~43%]로 확정(사전 밴드 20~48% 착지), 진술(카피 선택)도 행동과 같은 방향(매실청 우세). 세그 반응은 방향만 관측, 확증 없이 정직 종료
후속 검증(게이트 C)	1차 신호가 한 자릿수 → 효과크기가 아닌 방향 판정 . 예산상 유의성 검정은 불가(사전 인지)

(위는 3초 요약이다. 아래는 그 근거와 해석이다.)

1. 이 실험이 답하려 한 질문

본선에서 완벽하게 준비한 7분 PT를 수행했으나 탈락했다(심사위원 질문 0, 공식 피드백 0). “본선에 오른 것”은 성공이 아니다 — 정확한 사실은 “완벽한 발표에도 탈락”이며, 이 구분이 진단 방향 전체를 결정한다. 발표 완성도가 문제가 아니라면 남는 것은 *내용의 논리*이고, 나는 하나의 가설을 세웠다 —

H2 (탈락 원인 가설): 나는 차별점으로 매실청(향신료 재해석, A)을 전면에 세웠는데, 그 조준이 틀렸다. 완성도·외식대체(정공법, C)를 앞세웠어야 했다. (문서 B(의사결정 로그)의 ‘탈락 원인 가설(가)’와 동일한 가설이며, 문서 E(실행 기록)의 ‘강조점 판정(H2)’이 이를 판정한다.)

근거가 된 정황은 게이트 A(제품·광고비 투입 전의 무료 데스크 리서치)였다. 국내 쿠팡 리뷰 99건을 분석하니 개선 요구(갭)는 33%로 뚜렷했으나, 그 갭이 전부 완성도·외식대체(C)에 몰렸고 향신료 부담(고수)은 0%였다. “살아있는 수요 엔진은 C이고, 매실청은 소비자가 원한 과녁이 아니었다”는 의심이다.

이 가설을 감이 아니라 데이터로 판별하기 위해, 동일 제품·동일 타겟·동일 지면에서 강조점만 다른 **3-arm 가짜 상세페이지(fake-door: 실제 판매 없이 수요 신호만 측정하는 페이지)** 실험을 설계·집행했다.

arm	강조점	가설상 역할
T1	완성도·외식대체	내가 “옳았어야 했다”고 본 정공법(C)
T2	매실청·향 부담 해소	내가 “틀렸다”고 의심한 차별점(A)
C0	새우볶음밥 (더 싸고 간편)	의도적으로 가혹하게 잡은 대조군[尺]

(fake-door는 “출시 예정·결제 없음·개인정보 미수집”으로 정직하게 고지했고, 실제 브랜드명은 쓰지 않았다. 타이머·허위 할인 같은 조작 장치도 넣지 않았다.)

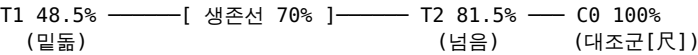
2. 결과 — 가설이 뒤집혔다

광고 기간·국가(한국)·고유 사용자로 정제한 GA 실측이다. 전환은 1차 신호(알림 신청 + 구매 클릭)를 누른 *고유 사용자* 수의 비율로 읽었다 — 같은 사람의 반복 클릭을 신호로 착각하지 않기 위함이다.

arm	강조점	유입 (pdp_view)	알림	구매	1차 신호 합	전환율	대조군 대비
T2	매실청	257	3	5	8	3.11%	81.5%
C0	(대조군)	157	3	3	6	3.82%	100%
T1	완성도	108	1	1	2	1.95%	48.5%

판정 기준(사전 등록 — 데이터를 보기 전에 잠근 것): 대조군의 **70% 이상**이면 생존, **35% 미만**이면 무용성으로 조기 종료.

생존선(70%) 기준으로 세 arm이 놓인 위치:



- **T1(완성도)**은 대조군의 **48.5%** — 생존선을 밀돌았다.
- **T2(매실청)**는 대조군의 **81.5%** — 생존선을 넘었다.
- **T2가 T1보다 명확히 우세하다** — 전환율 3.11% vs 1.85%(약 1.7배)이고, 실제 구매 클릭만 봐도 5명 vs 1명이다.
- 단, **T2도 대조군(100%)보다는 낮다.** 즉 매실청이 자[尺]를 *이긴* 것이 아니라, *자*에 **육박**했고 **완성도(T1)**는 *그 절반에 그쳤다*가 정확한 그림이다.

즉 내가 “틀린 과녁”이라 의심한 매실청이, 오히려 정공법보다 강한 신호를 냈다. **H2 가설은 반박됐다.** 문제는 매실청을 세운 것이 아니었다. 오히려 그것을 *충분히 세우지 않고* 완성도를 앞세운 조준이 약했을 가능성을 데이터가 가리킨다.

그렇다면 게이트 A(리뷰 99건, 고수 0%)는 왜 반대 방향을 가리켰는가. 두 데이터는 모순되지 않는다 — **표본이 달랐다.** 리뷰 코퍼스는 ‘이미 이 카테고리를 산 사람’만 담는다. 향신료가 부담스러운 사람은 애초에 동남아 간편식을 사지 않으므로, 그들의 목소리는 기구매자 후기에 0%로 나타난다(표본 선택 편향). 즉 ‘고수 0%’는 “A 수요가 없다”가 아니라 “A 타깃이 기구매자 집단에 부재한다”는 뜻이었다. 이 가능성은 게이트 A 단계에서 이미 인지하고 있었다 — 그래서 리뷰만으로 매실청을 기각하지 않고, 후기가 닿지 못하는 비구매자에게 까지 닿는 광고 실험(게이트 B)에 직접 판정을 위임해 두었던 것이다. 게이트 B는 그 위임에 위와 같이 답했다.

한 가지가 더 드러났다. T1만 봤을 때(48.5%)는 “정공법조차 대조군의 절반이니 제품 수요 자체가 약한가”로 읽힐 수 있었다. 그러나 T2를 켜자 같은 제품·같은 대조군에서 48.5%→81.5%로 뛰었다. **약한 것은 수요가 아니라 강조점이었다.** 이 구분은 T2를 켜지 않았다면 영영 보이지 않았을 것이다 — 그래서 “T1의 결과가 어땠든 T2는 반드시 측정한다”는 규칙을 설계에 미리 잠가둔 판단이 결과적으로 결론을 지켰다.

게이트 B가 남긴 공백 — 게이트 C(설문)가 짚 것

행동 실험은 방향을 줬지만 두 가지를 말해주지 않았다. 매실청이 겨는 수요층(향 기피)이 **인구에서 얼마나 되는지**, 그리고 사람들의 **진솔**이 이 행동과 같은 방향인지. 그래서 설문(게이트 C)을 사전등록 — 동어반복·프라이밍·사후 컷 조정을 수집 전에 차단 — 으로 설계해 파 일럿을 집행했다(유효 N=88, 총비용 ₩36,000).

- **수요층의 크기 — 확정.** 향 기피(협의 기준: “매우 그렇다”) 유병률 **33.0% [95% CI 24~43%]** — 사전에 잠근 목표 밴드 20~48%에 착지했고 모집 채널 간에도 강건했다. 본선에서 근거 없이 주장했던 “향 때문에 피하는 사람들”이, 실측으로는 **셋 중 하나**다.
- **진솔과 행동의 일치.** 두 소개 문구를 놓고 고르게 하자 매실청 카피가 우세했다(43% vs 27%) — 게이트 B의 행동 판정($T2 \geq T1$)과 같은 방향이다. 지갑의 방향과 말의 방향이 일치했다.
- **세그 반응은 방향까지만.** 향 기피층이 매실청 카피를 더 고르는 경향(+13%p)은 관측했으나, n=88의 검정력으로는 우연과 구분되지 않아($p \approx 0.18$) 설계에 사전 승인해둔 출구 그대로 “**방향 관측, 확증 없음**”으로 정직 종료했다. 게이트 B의 방향 판정과 같은 격이다.

3. 그렇다면 나는 왜 떨어졌는가 — 이 실험이 답한 것과 답하지 않은 것

먼저 두 개념을 분리해야 한다. **탈락 원인**(심사위원이 *실제*로 왜 떨어뜨렸나)과 내 자기진단(“매실청이 약한 무기여서 떨어졌을 것”이라는 내 추측)은 다르다. 공식 피드백이 0이었으므로 전자의 속내는 직접 확인할 수 없다. 이 실험이 검증한 것은 후자다.

	내용
이 실험이 답한 것	“매실청이 약한 무기여서”라는 내 자기진단은 틀렸다 — 매실청 강조(T2)가 완성도 강조(T1)보다 강한 신호를 냈다
답하지 않은 것	실제 탈락 원인. 이 실험만으로는 여전히 확정할 수 없다

그러므로 이 실험의 성과는 탈락의 *해답*이 아니라 **질문의 정련**이다. 가장 유력했던 용의자(“무기 선택이 틀렸다”)의 알리바이를 증명함으로써, 원인 후보를 ‘내용의 방향(무엇을 골랐나)’에서 ‘실행과 소구(어떻게 썼나)’로 이동시켰다.

원인 후보를 소구 축으로 좁히는 데는 근거가 하나 더 있다. 나머지 축 — 기술 현실성·가격·사업성 — 은 본선에서 부록까지 동원해 두껍게 방어했으므로(손익 구조·원가율·냉동 기술·로드맵을 각각 별도 부록으로 다뤘다), 그 부실을 탈락 원인으로 보긴 어렵다. 따라서 ‘PT에 문제가 있었다면’을 전제할 때, 남은 유력 후보는 강조점·소구 축이다.

그리고 대본을 다시 보면 결정적 관찰이 나온다. 나는 매실청을 단독 주인공으로 **전면에 세우지 않았다**. 매실청은 세 시나리오(향신료 회피 A·접근성 부족 B·기존 제품 불만 C)의 수렴 구조에서 A에 대한 답으로 다뤄졌고(“어떤 시나리오든 결론은 같다”), K-패치 매핑(고수→미나리, 타마린드→매실청, 강황→울금, 피시소스→멸치액젓)의 네 대체 중 하나였다. 더욱이 실무자 피드백을 반영하며 “단일 향신료 기피 논리”를 “3시나리오 양면 전략”으로 바꿨다 — 향신료(매실청) 축을 오히려 *희석*하는 방향이다.

그런데 실험은 그 완성도(C) 축이 오히려 **약한 축**임을 보였다. 내 최종 PT의 소비자 가치는 “부드러운 새콤함 · 5분 완조리 · 프리미엄 품질” 셋이었다. 첫째만 매실청(A) 축이고, 뒤의 둘(완조리·프리미엄)은 실험에서 T1이 전면에 세웠던 편의·완성도 번들 — 즉 약한 신호가 확인된 쪽이다. 나는 **약한 쪽 소구를 전면**에 둘 **갈고**, 강한 축이자 서류를 통과시킨 축(**매실청**)을 하나로 줄인 셈이다. 앞세운 것과 물린 것이 정확히 뒤바뀌어 있었다.

이 둘을 겹쳐 읽으면 하나의 가능성이 떠오른다 —

피드백 반응은 *논리적 방어력*(어느 가정이 틀려도 무너지지 않는 구조)을 높였다. 그러나 *구매 소구력* 차원에서는, 그 과정에서

게다가 희석한 그 축은 *아무* 축이 아니었다 — 높은 경쟁률의 서류 전형을 통과시킨, 이미 한 번 검증된 강점이었다. 제출한 식품 제안 원문의 중심이 바로 이 K-패치(향신료를 한국 재료로 재해석) 표였고, 접근성·완성도는 부차적으로만 적혀 있었다. 세 조각이 한 점으로 모인다 —

검증된 강점을 스스로 무디게 만든 것 — 이것이 강조점 오판의 가장 구체적인 형태다.

	총위	어떻게	무엇을
소구 (전면)	뾰족하게		서류를 통과시킨 매실청 축 하나를 앞세운다
방어 (후단)	두껍게		접근성·완성도를 그 뒤를 받치는 근거로 배치한다

게이트 C의 자유 응답은 이 ‘소구’ 축이 실제로 얼마나 섬세한 문제인지 실물로 보여줬다. 소비자가 카페에서 걸려 넘어진 최대 지점은 매 실청이 아니라 “타마린드”라는 단어였고(모른다·생소하다는 반응이 다수), ‘진짜 팟타이 맛이 아닐 것’이라는 정통성 우려는 향 기피층(7%)이 아니라 비기피층(31%)의 장벽이었다. 매실청 소구는 기피층에겐 다리, 정통파에겐 감점인 양날이다. 차별점의 **내용**이 옳다는 것(게이트 B)과 그것을 **말하는 방식**이 옳다는 것은 별개의 문제이며 — “소구는 뾰족하게”라는 교훈은 “누구에게, 어떤 언어로”까지 내려가야 완성된다.

4. 이 결론의 한계 — 정직하게

- 판정을 떠받치는 1차 신호는 한 자릿수다(T2 8명·T1 2명·C0 6명).
- 그래서 배수(1.7배)의 정밀도는 낮고 신뢰구간은 넓다 — 예컨대 **T1이 단 1명만 더 늘렸어도 전환율이 1.85%→2.78%(대조군의 73%)로 올라 생존선을 넘었을 것이다**. 한 자릿수 신호는 이만큼 출렁인다.
- 따라서 이 실험이 낼 수 있는 결론의 최대치는 **효과크기 추정**이 아니라 **방향** 판정이다.
- 대조 커피의 완성도 차이도 배수에 일부 섞였을 수 있다 — T1(완성도)은 C 축 단일 강조로 설계됐고 3시나리오 병렬이 아니었으나, T1 헤드라인이 T2만큼 날카로운 소비자 통점(pain point)을 짚지는 못했다. 다만 이는 완성도(C) 축이 본질적으로 감정적 통점이 약한 데서 주로 기인하는 것으로 보이며, 방향(A>C)은 유지되되 정확한 배수는 '축의 차이'와 '커피 차이'가 함께 작용한 값으로 읽어야 한다.

수집 방식에 대한 단서도 남긴다. T1(6/25~7/3)과 T2(7/4~7/6)는 순차로 수집했다. 서로 다른 시간대 수집은 시간 교란(그 사이 유입 성향이 달라져 비교가 오염되는 것) 우려가 있으나, 동일 타점·동일 지면·연속 평일 구간이므로 그 영향은 무시 가능한 수준으로 판단했다. — 비교 지표가 절대량이 아니라 전환율(비율)이라, 유입량 변동은 자동으로 통제되기 때문이다. 이 판단의 근거를 기록으로 남긴다.

5. 이 프로젝트가 실제로 증명하는 것

핵심은, 내가 옳았음을 증명한 프로젝트가 아니라 내가 틀렸음을 스스로 밝혀낸 프로젝트라는 것이다. 나는 탈락 후 “매실청을 잘못된 과녁으로 쏘었다”는 가설을 세웠고, 그것을 반증할 수 있는 실험을 직접 설계해 진행했으며, 그 결과 내 가설이 틀렸음을 데이터로 확인하고 결론으로 받아들였다. 자기 아이디어가 옳았다는 결말보다, 자기 진단이 틀렸음을 스스로 데이터로 뒤집는 결말이 드물고 또 어렵다. PM에게 필요한 것은 자기 판단을 받아들이는 능력이 아니라, 그것이 틀렸을 때 데이터 앞에서 인정하고 방향을 바꾸는 능력이라고 믿는다. 이 프로젝트의 정점은 그 전환에 있다.

그리고 그 전환은 우연이 아니라 설계의 결과였다. 반증이 가능하려면, 반증이 가능하도록 실험이 짜여 있어야 했다 —

- 완벽한 발표에도 탈락한 사건을, 감정이 아니라 검증 가능한 가설(H2)로 좁혔다.
- 자기 가설에 유리하지 않은 구조를 일부러 택했다 — 의도적으로 가혹한 대조군, 사전 등록된 판정 기준, “T1의 결과가 어쨌든 T2는 반드시 측정한다”는 게이트. 이 규칙이 없었다면 T1이 생존선을 밀돈 시점(48.5%)에서 ‘정공법도 안 통한다’며 실험을 닫았을 수 있고, 그랬다면 매실청이 강조점이었다는 결론은 관측되지 않았을 것이다.
- 광고를 실제로 집행하고, 계측 파이프라인을 봇·중복·기간 오염으로부터 정제했다 — 이벤트 수가 아니라 고유 사용자로, 광고 기간·국가로 칼같이 잘라, 전환을 표면 지표가 아닌 실측으로 읽었다.
- 표본의 한계를 알고 설계했고, 결론을 그 한계 안에서만(효과크기가 아니라 방향으로만) 주장했다.
- 후속 검증(게이트 C)에서도 같은 규율을 반복했다 — 분석 규칙을 수집 전에 잠갔고, 확증하지 못한 것(세그 반응)을 확증했다고 쓰지 않았으며, 사전 승인해둔 출구(“본조사 미검증, 방향만”)로 종료했다. 규율이 한 번의 사건이 아니라 반복 가능한 방식임을 보였다.

즉 이 프로젝트는 두 가지를 동시에 보여준다. 하나는 자기 진단을 데이터로 뒤집는 **지적 정직성**이고, 다른 하나는 그 뒤집힘이 관측될 수 있도록 실험을 설계·집행·해석하는 **실행 역량**이다. 전자가 이 케이스의 결론이라면, 후자는 그 결론이 신뢰할 만한 이유다.

판정 근거: GA4 탐색(광고 기간 국가 한국·고유 사용자 기준), 3-arm fake-door, 사전 등록 판정 기준(생존 70%·무용성 35%). 1차 신호는 알림 신청·구매 클릭을 누른 고유 사용자. 표본 한계상 결론은 방향 판정으로 한정한다. 상세 수차·계측 규율은 문서 E(게이트 B 실행기록) §9, 게이트 C(설문)의 사전등록·집행·판정은 E §9.4와 게이트C_파일럿_분석보고서에, 가설 형성 근거는 문서 D(게이트 A)·B(의사결정 로그)에 있다.