WANDER · COMPANION DOCUMENT B

Wander: Research & Evidence Companion

How I tested Wander, what the evidence supports, and where it stops.

AUTHOR	Shawn Farnum
PRODUCT	wandr.it · publicly available · v1.0 in progress
CASE STUDY	shawnfarnum.com/work/wander
EVIDENCE THROUGH	November 4, 2026
VERSION	1 · November 4, 2026

Every figure here is an aggregate. This document contains no participant-level records, participant identifiers, contact details, consent submissions or quotations.

The studies are small and self-selected. Read the counts as formative signals, not population estimates.

Contents

00 How to read this document

01 Research objectives

02 Alpha01: initial pilot

03 Firstvisit01: first visit, unmoderated

04 Study limitations

05 Design-decision records

06 Technical evaluation

07 Experiments and dispositions

08 Architecture reference

09 Versions and evidence dates

10 What remains unverified

00

How to read this document

Wander's evidence comes in three kinds, and I keep them apart because they support very different claims:

EVIDENCE LEVELS USED THROUGHOUT

Level	Means	Doesn't mean
Production	Part of the publicly released product (v0.7.8), or recorded in a study run on it.	Proven effective, or continuously available.
Local candidate	Built and tested on my machine after the release.	Deployed, or verified on the hosted product.
Exploration	An idea or open question.	Planned, built or validated.

01

Research objectives

- Does Wander consistently show people things worth discovering, and does that make them want one more Wander?
- Can wandering itself introduce the product, without my explanation or interface instructions?
- Do the later concepts (Trail, How Far?, reactions, Wander From Here) make sense when they appear?
- Does the topic metadata behind selection represent meaningful relationships between Discoveries?

The first three are questions for people. The fourth is a technical question, and section 6 explains why it still needed people to answer it.

02

Alpha01: initial pilot

Design

Before the pilot, I wrote down a behavioral hypothesis: someone clicks Wander ten times and wants to click it an eleventh. Five real, pseudonymous participants consented, each with one session, followed by an optional debrief that all five completed. Research recorded only consented actions inside Wander. Opening a destination records a click, not time spent or activity on the external site. My own use of the product was excluded.

ALPHA01 INITIAL WAVE, FROZEN AT FIVE PARTICIPANTS (EXPORT OF SEPTEMBER 27, 2026)

Measure	Result
Participants / sessions	5 / 5
Reported finding something they probably wouldn't have found alone	5 of 5
Continued from Discovery 10 to 11	5 of 5
Continued from Discovery 15 to 16 (of those reaching 15)	4 of 4
Reached Discovery 1 / 5 / 10 / 15 / 20	5 / 5 / 5 / 4 / 2
Discoveries shown, total	132
Discoveries per participant (median; range)	19; 14 to 60
Destinations opened (share of impressions)	32 (24.2%)
Participants opening at least one destination	5 of 5
Reactions: Interesting / More like this / Not for me	40 / 26 / 11
Participants who changed distance (changes)	3 (11)
Saves; Wander From Here uses	0; 0
Return sessions	0

What it supports

- People could use the core Wander action without instruction.
- In this small sample, the loop sustained continuation well past ten Discoveries.
- All five reported a worthwhile, unexpected find, and all five opened at least one destination.

What it doesn't

- Retention, satisfaction, statistical significance or product-market fit. There were no return sessions and no comparison group.
- The value of Save or Wander From Here, the right distance controls, corpus size or ranking weights.
- A uniform mental model: people described Trail and the distance controls in different ways, and some weren't sure what they did.

One participant contributed 60 of the 132 presentations and 57 of 77 reactions, so pooled activity leans heavily on one person. One participant needed help finding the study's Finish control; the data can't separate their subsequent behavior from the rest. Explore is the default mode, so mode totals show the active mode, not necessarily a deliberate choice.

03

Firstvisit01: first visit, unmoderated

Design

Firstvisit01 tested the premise that wandering itself could introduce the product, without me present. Participation was remote and unmoderated, on the Production site. Consent came before exploration, and an optional questionnaire was offered after the fifth Discovery. The study build was prepared on October 21, 2026, and the study closed on October 27, 2026.

FIRSTVISIT01, VERIFIED ANALYSIS EXTRACT OF OCTOBER 26, 2026

Measure	Result
Consented participant records	11
Started exploring	10
Reached five Discoveries	5
Questionnaires submitted	3
Of those, submitted after recorded exploration	2
Median recorded trail length	3 Discoveries
Presentations, total; from the most active record	103; 60
Records whose last action involved study controls	7

Interpretation

Several people began without guidance, which is consistent with the premise. But exploration was short and uneven, and with three questionnaires, basic comprehension and perceived value remain unresolved. Seven records ending on study controls is an association; it doesn't tell me why anyone stopped. A destination open

shows someone went through the door, not that they read or enjoyed what was there.

I didn't respond to this study by adding interface instructions. I want to understand what participants meant before changing the experience.

Eleven records don't equal eleven independently verified unique people, and the sample was self-selected. Qualitative coding of the questionnaire answers is awaiting my own review, so no coded themes are reported here.

04

Study limitations

- **Small, self-selected samples.** Five participants in alpha01's initial wave; 11 records in firstvisit01.
- **Concentrated activity.** In both studies, one participant or record produced roughly half or more of all presentations.
- **Behavior inside Wander only.** Opens are clicks. There's no measure of time on, or enjoyment of, the destination.
- **No return or comparison data.** Neither study establishes sustained voluntary return, and neither compared Wander with an alternative.
- **Study wrapper effects.** One alpha01 participant was helped with the Finish control, and seven firstvisit01 records ended on study controls.
- **Zero use is ambiguous.** A feature nobody used could be hidden, unclear or unnecessary. Short first sessions can't tell those apart.

05

Design-decision records

DDR 1 · Wandering is the onboarding

CONTEXT	A discovery product can't ask for tastes before someone has experienced it.
DECISION	One-action first screen. Trail, How Far?, reactions and Wander From Here appear after Discoveries 1, 2, 3 and 4. An account is offered only after real history.
EVIDENCE	Product intent, implemented in the release. Alpha01 participants used the core action without instruction; firstvisit01 left comprehension unresolved.
STATUS	PRODUCTION Hypothesis still open.

DDR 2 · History belongs to the identity, not the trail

CONTEXT	"Never show me that as new again" had to survive new trails, reloads and account creation.
DECISION	The server filters out everything an identity has seen. A skip counts as seen but not disliked. When nothing unseen remains, Wander says so.
EVIDENCE	Enforced in the API and database.
STATUS	PRODUCTION

DDR 3 · Accounts are optional and only preserve

CONTEXT	A sign-up gate would contradict the one-action premise.
DECISION	Anonymous start; accounts carry history across devices; export and deletion under Your data; research consent separate from ranking.
EVIDENCE	Implementation.
STATUS	PRODUCTION

DDR 4 · Static previews, reviewed destinations

CONTEXT	Loading arbitrary third-party pages inside Wander would put their scripts in Wander's page.
DECISION	Static preview images; destination URLs treated as untrusted and verified before editorial approval.
EVIDENCE	Implementation. Verification at ingestion doesn't guarantee a destination stays safe or available.
STATUS	PRODUCTION

DDR 5 · Keep wandering stays reachable on short phones

CONTEXT	On short screens, the preview and description pushed the primary action below the first screen.
DECISION	Compared tighter spacing, a shorter preview and moving the action earlier. Kept the preview; moved Keep wandering closer to the destination link.
EVIDENCE	Measured comparison of matched local layouts at 375 by 667. Not a participant preference test.
STATUS	DECIDED

DDR 6 · How Far? doesn't go above the primary action

CONTEXT	Placing the modes above Keep wandering made their relationship to the next Discovery explicit.
DECISION	Rejected: it pushed the primary action lower in three of four matched short journeys and added three keyboard stops. A disclosure menu would hide the choices. A compact after-action variant is frozen for comparison.
EVIDENCE	Local layout and keyboard measurement. No participant comparison result exists.
STATUS	LOCAL CANDIDATE Not a winning design.

DDR 7 · Fix defects; hold open questions for research

CONTEXT	AI-assisted building made it cheap to answer every unclear finding with a feature.
DECISION	Fix measurable reachability problems. Don't redesign Trail, How Far?, Save or Wander From Here from a five-person pilot, and don't add instructions after firstvisit01.
EVIDENCE	Alpha01 and firstvisit01 limits (sections 2 to 4).
STATUS	STANDING RULE

DDR 8 · Close failed experiments instead of tuning them

CONTEXT	It's tempting to keep adjusting an alternative until it passes.
DECISION	Set gates before running; close an experiment that fails its gate.
EVIDENCE	Section 7.
STATUS	STANDING RULE

06

Technical evaluation

Method: an offline topic graph

To check whether the metadata supports meaningful journeys, I modeled the corpus as a graph: each Discovery a node, linked to every Discovery it shares a topic with. I measured two versions. The any-topic graph includes nine broad family topics (such as art, history and play) that ingestion appends by pattern, which is what the relevance score sees. The specific-topic graph excludes them, which makes it the stricter approximation, since a reviewer didn't choose those links. No clustering algorithm runs; components are simply what's reachable.

OFFLINE SNAPSHOT, 118 PUBLISHED DISCOVERIES, 6,903 POSSIBLE PAIRS (HISTORICAL, NOT THE CURRENT LIVE CORPUS)

Pair relationship	Pairs	Share
Share a specific topic	239	3.5%
Connected only through family topics	2,914	42.2%
Share no topic	3,750	54.3%

On specific topics alone, the graph splits into 39 components. The largest holds 51 Discoveries (43.2%), and 31 Discoveries are isolated, 30 of which connect only through family topics. With family topics included, 117 of 118 sit in one component. Broad topics do most of that work: art is the only link for 902 pairs, history for 822.

These numbers describe metadata connectivity. They aren't measures of user value, and an isolated Discovery can be excellent. Nothing was removed, retagged or reranked because of them.

Method: a single-trail inspector

An offline inspector explains one trail at a time: each move's path in both graphs, whether it crosses components or depends on a bridge, and flags such as family-topic-only connections or Explore and Wild collapsing to the same choice. On one simulated trail, no selected move had a specific-topic path, and all 11 moves changed component. That describes one simulated trail, not people.

Method: blind human transition ratings

Because structural tidiness isn't evidence of a good transition, a small registered review asked people to rate constrained sets of transitions without knowing which condition produced them. Specific-topic transitions were rated as more related in a limited set of cases. Other findings were mixed or unsupported. That doesn't establish that topic proximity should decide ranking, and it covers constrained transitions, not open-ended journeys.

Method: gated engine comparisons

The selection engine can take injected random sources, so an offline harness can compare two variants on the same random draws. Production passes none. Each alternative in section 7 had quality gates written down before it ran.

07

Experiments and dispositions

RECOMMENDATION EXPERIMENTS

Experiment	Outcome	Disposition
Context-aware Explore shortlist	Increased the candidate pool; fell below its quality comparison; favored sparsely tagged destinations.	Not adopted

Experiment	Outcome	Disposition
One-over-rank draw	Failed quality gates across modes.	Not adopted
Eligibility filter	Passed its hard gates; doesn't improve variety by design.	Not integrated
Challenger E3	Failed its preregistered technical gate: overall source-balanced quality change missed the required bound; ten of twelve crossed cells failed.	Stopped; later phases not run

None of these entered the app, and none should be read as nearly successful. They also don't validate the current engine. They show that changing weights or draws didn't produce a better-evidenced alternative.

08

Architecture reference

The components involved in a Wander, as of the v0.7.8 release:

COMPONENTS

Component	Responsibility
Web client	The interface, the teaching sequence and the dialogs. Sends one action per request.
Wander API	Receives actions (wander, wander from here, new trail, reaction, save, dismiss offer) and returns the updated state.
Wander operation	Framework-independent request logic shared by every client and the database tests: identity, teaching moments and the account offer.
Discovery engine	Eligibility, scoring, the mode weights, the top-three draw and Wander From Here's starting-point topics.
Store (Postgres)	Identity-wide history, trails, reactions and saves.
Curated corpus	The versioned, reviewed Discovery index the application loads.
Ingestion	Treats candidate URLs as untrusted and verifies them before editorial approval; appends family topics.
Offline evaluation	The topic-graph audit, the trail inspector and the gated engine comparisons. Never in the request path.

09

Versions and evidence dates

TIMELINE OF THE EVIDENCE IN THIS DOCUMENT

Date (2026)	Item	Level
September	Project begins	Production
September 27	Alpha01 initial-wave export (frozen at five participants)	Production
October 21	Firstvisit01 consent and questionnaire build prepared	Production
October 26	Firstvisit01 verified analysis extract	Production
October 27	Firstvisit01 closes; v0.7.8 released (last documented release)	Production
October 28	Accessibility correction, corpus trust and reliability work	Local candidate
November 4	Public availability checked; this document	Production

The 118-Discovery graph audit is a historical snapshot, separate from today's live corpus. A point-in-time availability check isn't evidence of uptime.

10

What remains unverified

- **Firstvisit01's final extract.** Operator-reported totals at closing matched the October 26 extract, but I haven't verified a final participant-level extract after closure.
- **Qualitative themes.** Coding of questionnaire answers is awaiting my review, so it's left out.
- **Accessibility on real setups.** The dialog fix passed local probes and automated checks. Physical devices, assistive technology and the hosted product haven't been tested. It isn't a certification.
- **Local reliability and corpus work.** Tested locally, not deployed.
- **The How Far? comparison.** The compact variant is frozen; no participant result exists.
- **Recommendation quality.** No experiment established a better production approach, and the current one isn't validated.
- **Outcomes.** No evidence yet of sustained return, broad appeal, retention, conversion or revenue.
- **Beyond the web.** Online and offline discovery is an exploration. Nothing is built or tested.