
What the Score Doesn't Tell You

A behavioral analysis of 298,553 Steam reviews

Every game on Steam wears a single number — **85% positive**, say — and we read it as a verdict. This report takes that number apart. Across a quarter-million reviews, the same headline score turns out to hide *when* a player reviewed, *whether they stayed*, *how they wrote*, and *who they were*. Four findings, each tested the same demanding way, each a layer the score quietly compresses away.

THE REFUND WINDOW

Reviews written before the two-hour refund deadline are far harsher — a cliff right at the line.

A REVIEW IS A GOODBYE

Pan a game and you stop playing it; recommend one and you keep going. The score encodes which.

NEGATIVITY IS VERBOSE

Negative reviews run more than twice as long. Brevity, it turns out, is a positive act.

THE VETERAN IS HARsher

The same game scores lower from players with bigger libraries. Experience raises the bar.

The number everyone trusts

A Steam rating looks like a verdict: a single percentage, the crowd's thumbs up or down, distilled. It is one of the most consulted numbers in gaming — players buy or skip on the strength of it. But a percentage is a *compression*. It flattens hundreds of thousands of individual decisions, each made by a particular person at a particular moment, into one digit of signal. This report decompresses it.

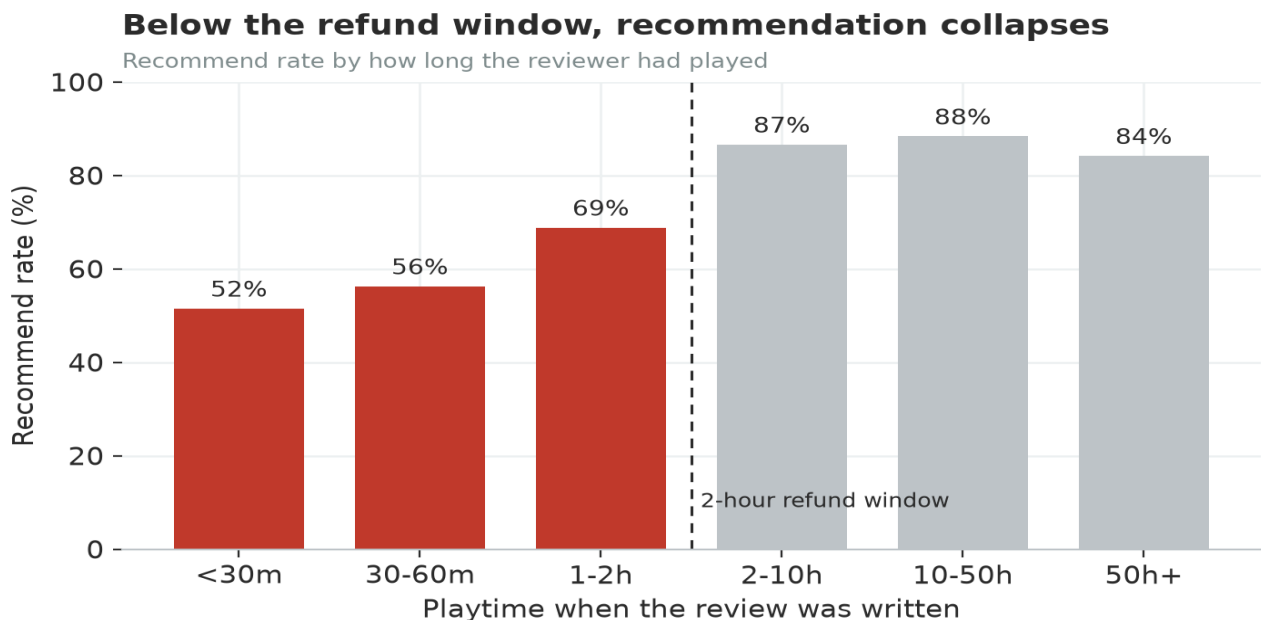
The data is 298,553 reviews across 50 games, in 30 languages — collected from Steam's public API, cleaned, and analysed not in the aggregate but *within each game*, so that a pattern has to prove itself title by title before we believe it. That discipline is the spine of everything here: it is the difference between a real effect and a trick of which games happened to land in the sample. Four patterns survived it. Here is the short version.

What we looked at	What the score was hiding
When the review was written	Below the 2-hour refund line, recommendation falls to ~61% — against ~86% after. A cliff at the deadline.
Whether the player stayed	Recommenders play a median 3h more after reviewing; panners play ~0h. 63% of negative reviewers never return.
How the review was written	Negative reviews run 83 characters to positives' 35 — more than twice as long. Brevity signals approval.
Who was reviewing	Players with 200+ games recommend at ~79%, the smallest libraries at ~93%. The same game, a different verdict.

None of these is visible in the rating itself. Each is a behaviour the number averages into silence. The rest of this report takes them one at a time — moving from the mechanical to the human: from a refund clock, to the moment a player walks away, to the shape of what they write, to who they are when they write it.

1 · When — the refund window

It is late, you have owned the game ninety minutes, and it is not clicking. Steam will refund anything under two hours played — so you have a decision to make, and a deadline to make it by. Quietly take the money back, or leave a warning on the way out. That small, ordinary moment turns out to leave a mark across the entire dataset.



Recommend rate climbs steeply through the sub-two-hour buckets, then locks flat. 298,553 reviews across 50 games.

Sort reviewers by how long they had played when they wrote, and the pattern is stark. A reviewer who played under thirty minutes recommends the game barely half the time. Cross the two-hour line and the rate jumps to a steady ~86% — and then stays there, flat, whether the player logs ten more hours or a thousand. The score doesn't drift with enjoyment; it steps, sharply, right at the refund boundary.

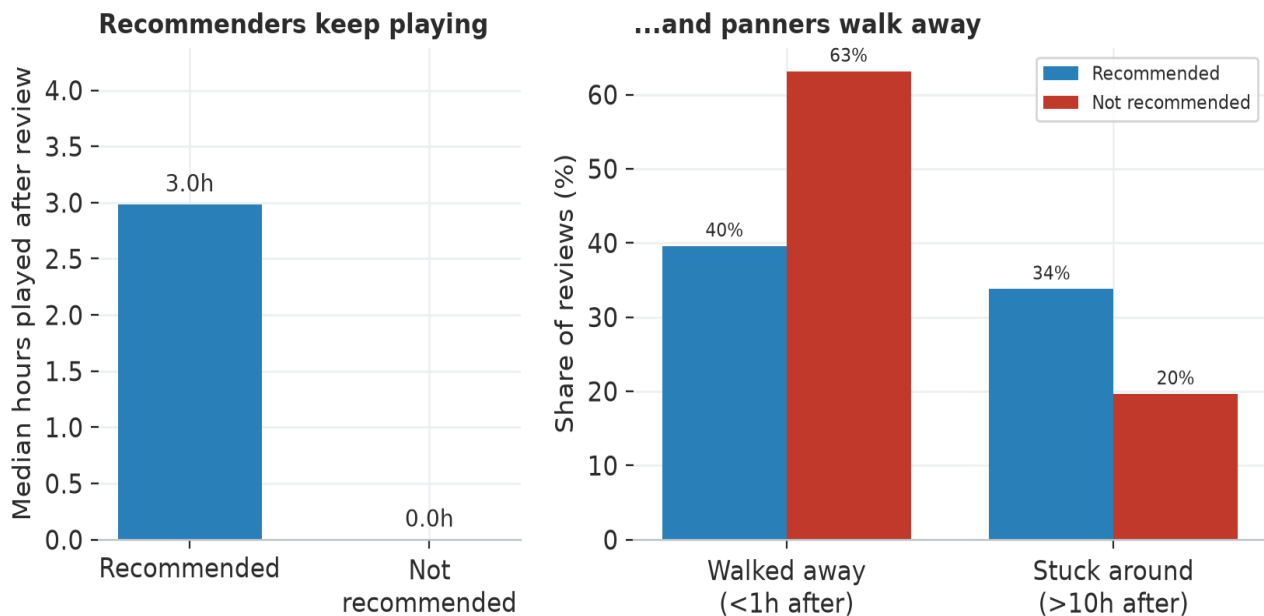
And this isn't a couple of angry games dragging an average around. The pattern holds *within* almost every individual game in the set, and an entirely separate signal — Steam's own flag for reviews that were actually refunded — points the same way: those reviewers cluster inside the window and recommend at just ~20%. A multivariate model agrees, with playtime the strongest predictor of a positive review even after the game, the review's length, and its votes are all controlled for.

Reviewers who write before the refund window closes are far harsher than those who write after.

What we can't say for certain is *why*. Perhaps the deadline itself manufactures angry reviews — play briefly, warn the world, reclaim your money. Or perhaps players who were never going to like the game simply quit early, so short playtime and negativity share a cause rather than one driving the other. Observational data can't fully separate the two. But the sharpness of the cliff — a clean break at exactly two hours, not a gentle slope — leans toward the window doing real work. Whatever the hand, the fingerprint is unmistakable.

2 · Whether they stayed — the goodbye

A review reads like a verdict handed down from outside the game. But it is written from *inside* a relationship with it — at some particular moment, with more of that relationship still to come, or not. Because the data records both a reviewer's total playtime and their playtime at the moment they wrote, we can see what they did next. The answer is quietly poignant.



Left: median hours played after the review. Right: who walked away versus who stuck around.

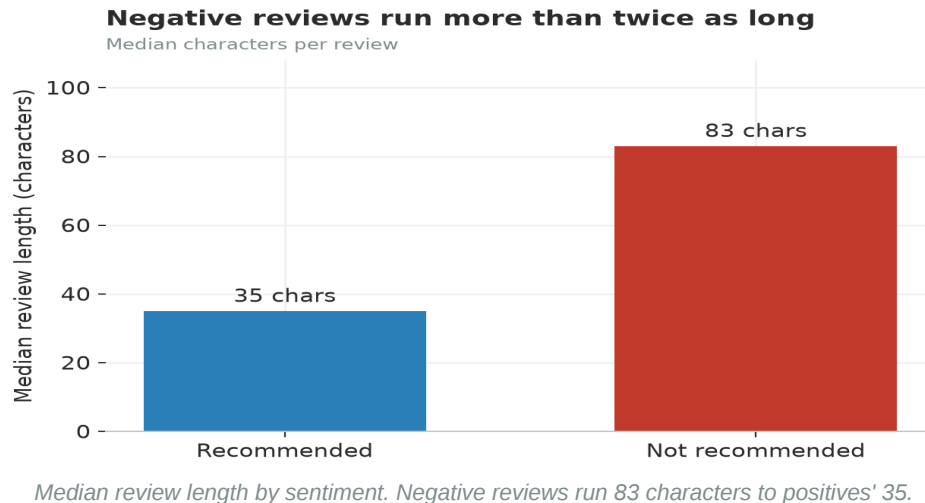
Players who recommend a game keep playing it — a median of 3 more hours after they review. Players who pan it mostly stop: their median additional playtime is essentially zero. Around 63% of negative reviewers never meaningfully return to the game after reviewing it. The negative review isn't a pause for feedback partway through; it is one of the last things the player does before leaving.

A recommendation is a love letter written mid-relationship. A pan is a goodbye note on the way out.

The striking part is that words and behaviour agree, with no need to read a single sentence of the review. The sentiment a player typed predicts what they physically did next — keep playing, or close the game for good. And like everything here, it holds title by title, not just in the pile: inside almost every game, its recommenders go on to play longer than its detractors. The score, it turns out, quietly encodes whether the reviewer stayed.

3 · How they wrote — negativity is verbose

Think of the last time something delighted you, and the last time something let you down. Which reaction took more words? On Steam the answer is not close — and you can see it before reading a single review, in nothing but their length.



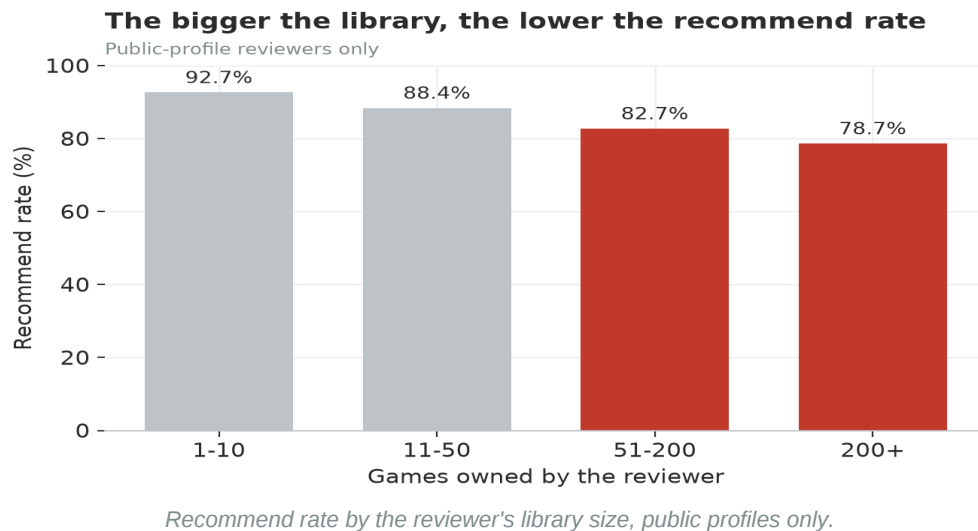
A positive review has a median length of 35 characters — about the size of “great game, highly recommend.” A negative one runs 83, more than twice as long. The effect is statistically overwhelming, and it holds within nearly every game. But the interesting part is *where* the gap lives. It isn't that a few furious players write enormous screeds — the median ignores those. It's the opposite end: the one-line review, the bare “10/10,” the curt “great game,” is almost always a recommendation. Over nine in ten ultra-short reviews are positive.

A positive review can be a reflex. A negative one tends to be an argument.

Why disapproval reaches for more words is a genuine open question — and an honest report should leave it open. Maybe complaint feels a burden of proof that praise never has to carry: “I loved it” is a complete review, while “I hated it” invites the question *why*. Maybe anger is simply more verbose, or maybe complaints name specific failures that take room to describe. Telling those apart would mean reading the words themselves, not just counting them — which is exactly where this project goes next. For now the asymmetry stands on its own, worth sitting with.

4 · Who they were — the veteran is harsher

The first three findings were about the review. This one is about the reviewer. Steam attaches to each review a rough measure of experience — how many games the author owns — and experience, it turns out, shows up in the scores. This is the gentlest of the four findings, and it comes wrapped in honest caveats, so it is offered plainly rather than oversold.



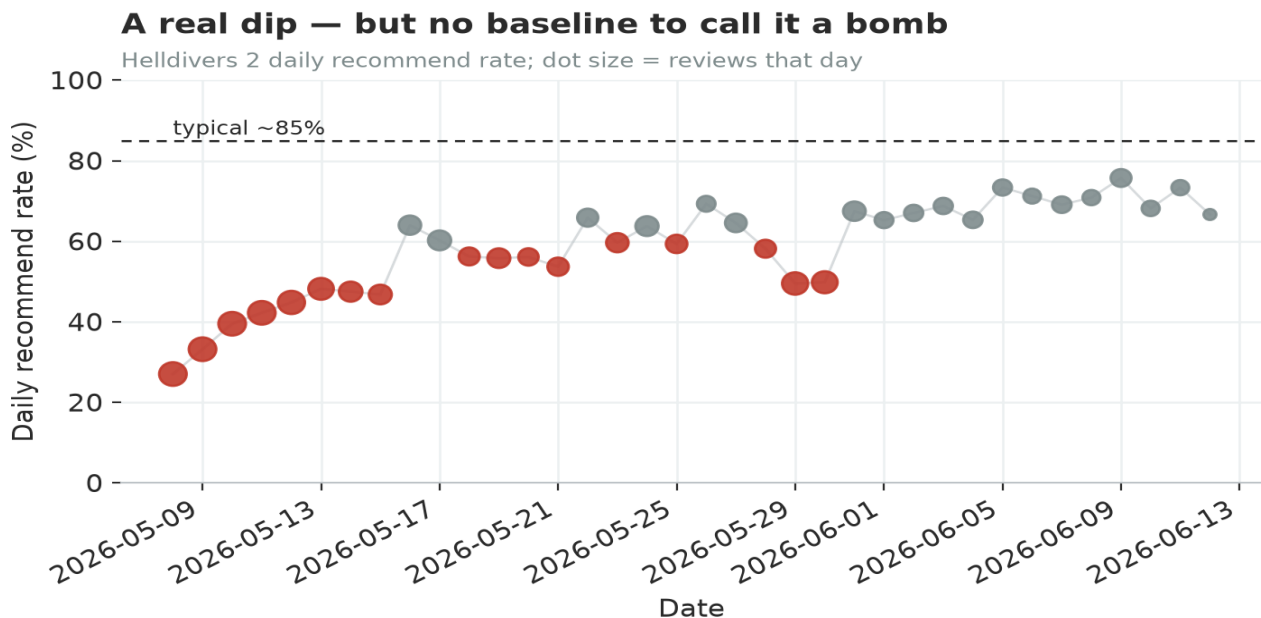
Among reviewers whose profiles are public, recommendation falls cleanly as libraries grow — from the ~93% of the smallest collections to ~79% for those owning two hundred games or more. The natural worry is that this is just selection: a player rich in games and time tries everything, including the dross, while a ten-game owner buys only sure things. If that were the whole story, though, the effect would vanish once we compare reviewers of the *same* game — and it doesn't. Title by title, the player with hundreds of games behind them is harder to please.

Experience raises the bar. The more of a medium you've seen, the less easily a new entry impresses you.

It is a familiar truth — the seasoned film critic, the lifelong reader, harder to win over than the newcomer — now visible in a quarter-million game reviews. The same game earns a different verdict depending on who is holding the controller. Which is, in the end, the whole point of this report: the score is not a property of the game alone.

The things we didn't claim

A report is only as trustworthy as the findings it was willing to throw away. We tested roughly a dozen ideas; four survived. The rest were set aside on purpose, and the most interesting of them is worth showing — because it looks like a finding and isn't.



Helldivers 2's daily recommend rate climbs out of a deep dip — but our window begins near its bottom.

For two days in May 2026, Helldivers 2's daily recommend rate sat near 27%, far below its usual ~85%, before climbing back over the following month. It has every appearance of a review bomb. And yet we decline to call it one — because a review bomb is a claim about *anomaly*, and to prove a spike is unusual you need the game's ordinary history to compare against. Our collection reaches back only weeks; it captures the recovery but never the calm before. We can show the dip. We cannot prove how strange it was. So we say the smaller, true thing, and leave the larger claim alone.

Several other tempting patterns met the same fate. Pricier games looked slightly worse-reviewed, but on only forty-odd titles the signal was too fragile to trust. Critics and players turned out to broadly *agree* — and agreement is no story. Different language communities showed different rates, but tangled inextricably with which games they play. Free copies produced almost exactly the same sentiment as paid ones — a clean null. And Steam's built-in “helpfulness” score sat at its default value for three-quarters of all reviews, hollow at the core. None became a chapter. Naming them is what earns the four that did.

What the score really is

Set the four findings side by side and a single idea emerges. The headline rating is not a verdict on a game so much as a *compression* of human behaviour — of when people reviewed, whether they stayed, how they wrote, and who they were. Decompress it and the structure underneath is legible, consistent, and often surprising: a refund clock bending sentiment, a goodbye encoded in playtime, approval hiding in brevity, a bar that rises with experience.

Two honest limits point the way forward. Several questions went unanswered for want of *history* — our window is recent by design, which a fuller crawl would extend. And the deepest question of all — *why* a negative review reaches for more words — went unanswered because we deliberately studied the *shape* of reviews, not their *content*. We counted words; we never read them. That is the next phase of this project: turning from structure to language, and asking how much the words reveal that the numbers around them cannot.

What the score doesn't tell you, this report found in the structure around it. The rest is in the words.

Methods, source code, and the full analysis notebooks: github.com/arda-basarici/steam-reviews · Findings are validated within individual games; all figures regenerated from the processed dataset.