



MathClash — Product Thesis

One claim, deliberately narrow. Everything we build serves it; everything that doesn't, we cut.

Drafted May 2026 · companion to the MathClash demo, the Build Plan (docs/MathClash-Build-Plan), and the research brief (docs/MathClash-Research-Competition-K12)

THE THESIS

Near-peer, low-stakes, anonymous, repeated 1-versus-1 competition — wrapped around a real difficulty ladder — is one of the few motivational structures with *consistent* evidence behind it for **confident and high-achieving students. MathClash exists to deliver that structure as a product, and to be the **cheapest way to give a "stretch-and-rank" experience to capable kids whose schools will never provide one.****

If that sentence is true, MathClash should exist. If it's false, MathClash shouldn't — and the bets on page 5 are how we'll know which.

Every clause is load-bearing — here's why each one is in the thesis

CLAUSE	WHY IT'S THERE (AND WHAT WE'D BE WITHOUT IT)
Near-peer	Rank and social-comparison effects work through standing among <i>comparable</i> others; a hopeless or trivial mismatch motivates no one. Drop it and you get rage-quits and demoralization. → level/skill-matched duels.
Low-stakes	Competition splits people into "I want to win" vs. "I'm scared to lose"; low stakes keeps far more students in the first bucket. Drop it and you build an anxiety machine — exactly the failure mode parents and schools fear. → no grades, nothing permanent rides on a loss, easy off-ramps.
Anonymous	Wanting to <i>outperform</i> others helps performance; wanting to <i>not look dumb in front of people</i> hurts it. Anonymity strips the audience cue — and also defuses harassment of minors. Drop it and the motive curdles. → randomized handles for everyone, the user included.
Repeated	One-shot high-stakes contests select; a stream of small ones <i>develops</i> . The value compounds over many low-consequence matches, not one big one. → frequent short duels, no "tournament day" pressure.
1-versus-1	Competitive drive is highest with few rivals and falls as the field grows (the "N-effect"); 1v1 is the maximum. Drop it for a giant global ladder and you've diluted the core mechanic. → duels as the primitive; leaderboards are divisioned, not one list.
Real difficulty ladder	The payoff for capable kids is the <i>stretch</i> , not the trophy; an absolute standard ("you cleared the level-8 problem") is also what keeps the player who lost the match still motivated. Drop it and you're a worksheet with points. → progressively harder problems, worked solutions, points scaling with difficulty.
Confident / high-achieving students	This is the segment the evidence actually supports — and, not coincidentally, the one schools serve worst at scale. We are <i>not</i> claiming this helps the general K-12 population (for them, cooperation beats competition). Naming the segment keeps us honest and keeps the product focused.

IN ONE BREATH

There is a capable kid — two grade levels ahead in math, no gifted track, no math team, parents not plugged into the talent-search world — who is bored, under-stretched, and has no rivals. School can't or won't fix that at scale. The competition-elite platforms aren't aimed at her; the mainstream practice apps have no competitive or social layer; the gamified K-8 apps aren't built for stretch. MathClash is the product that walks into that gap with a structure the research backs for exactly this kid.

1 WHO THIS IS FOR – THE CAPABLE KID THE SYSTEM DOESN'T SERVE

Primary user: a student roughly grades 6–10 who is meaningfully ahead of their classroom in math, intrinsically interested, and **structurally underserved** — no separate track, no competition team within reach, no parent steering them to AoPS/CTY/Math Circle, often a school that doesn't (or can't) accelerate. They don't lack ability; they lack **stretch** and **rivals**. The day-to-day symptom is boredom: "math is the easy class," and there's no one to push against.

Why the system leaves them there: separating high achievers into their own classroom *does* work — there's quasi-experimental evidence that it produces large, durable gains — but most districts don't do it (cost, politics, logistics, equity optics done wrong). And the kids who *do* reach the top of math are heavily concentrated in a small set of advantaged schools; a lot of capable students — disproportionately girls — never get there. The talent is real and widespread; the on-ramps are scarce and clustered. (Evidence and citations: docs/MathClash-Research-Competition-K12, §2–§3.)

Secondary users (downstream, not v1): the parent who wants their kid challenged and can't get the school to do it; the math-club or accelerated-track teacher who needs something for the kid who finishes early; eventually a district that wants a "challenge layer" it can switch on for its high achievers without standing up a gifted program. These widen the funnel later — they are not the v1 wedge.

Explicit non-user: the median or struggling K-12 student. For them the evidence favors cooperative and individualized structures over competition; serving them is a different product with a different motivational design, and trying to be both is how the focus dies (see §4 and §8).

2 WHY THIS WEDGE IS OPEN – THE COMPETITIVE MAP

WHO'S THERE	WHAT THEY DO	WHY THE CAPABLE-BUT-UNSERVED KID STILL FALLS THROUGH
School itself	Tracking / gifted programs where they exist; math teams at some schools.	Most districts don't track; teams cluster at advantaged schools; the marginal capable kid in an ordinary school gets nothing.
AoPS, CTY, Math Circles	Rigorous content + community + contest prep for the <i>already-identified</i> competition elite.	You have to already know it exists and opt in; it's pitched at the top few percent who self-select; it isn't trying to <i>find</i> the unserved kid.
Khan Academy, DeltaMath, IXL	Curriculum-aligned practice at scale; mainstream grades 6–12.	Essentially no live, social, head-to-head <i>competitive</i> layer; built for coverage, not stretch-and-rank; no "you're in the top X%, here's where to go next."
Prodigy and similar	Gamified math "battles," mostly K–8.	Engagement-first, light on rigor and on real difficulty progression; not built to <i>stretch</i> a capable older student or to plug into the math-talent pipeline.

The white space, stated as a sentence: **competitive stretch — and discovery — for capable students whom the school doesn't track and the elite programs don't reach**. Nobody is squarely in it. (Strategic positioning and the build-vs-partner options: docs/MathClash-Build-Plan, §1 and §7–8.)

3 WHY COMPETITION — AND WHY *THIS* FLAVOR OF IT

The honest baseline from the research: competition has **no average effect on performance**; it splits people, helping those it pushes toward "I want to outperform" and hurting those it pushes toward "I'm scared to look bad." That sounds like a reason *not* to build on competition — until you notice that the students it reliably helps are **precisely our target segment**, and that the split is steerable by design. The thesis's qualifiers are not garnish; each one is a knob set to keep the experience in the half of the population (and the half of the psychology) where competition works. The mapping:

DESIGN KNOB (IN THE THESIS)	EVIDENCE IT'S TUNED TO	WHAT WE'D GET IF WE SET IT WRONG
1v1, small divisions CORE	Competitive motivation peaks with few rivals and falls as N grows (the N-effect).	One giant global leaderboard: a number most kids can't move, so most stop trying.
Level/skill-matched opponents	Rank and relative-feedback effects depend on standing among comparable peers; mismatch is inert or demoralizing.	A 7th grader matched into Calculus, or against a bot they can't lose to: no signal, no stretch, no fun.
Anonymous handles	"Normative" performance-approach goals (outperform others) help performance; "appearance" goals (look able to others) hurt it.	Real names + an audience: the motive flips toward fear of public embarrassment, and you've also opened a harassment surface on minors.
Low stakes; loss ≠ failure; reward improvement/streaks/accuracy, not just wins	Competition triggers both approach and avoidance motives; "losers" stay motivated when given an absolute standard and positive feedback rather than a bare W/L or a "you-lost-no-reward" mechanic.	High-stakes, rank-graded, winner-take-all: an anxiety machine that's worst for the most invested kids — the well-being failure mode.
Real difficulty ladder + worked solutions inside every match	For capable students the gain is the stretch, not the trophy; precocious students do best given genuinely harder work, not "grade level."	Same-difficulty problems with points sprinkled on: an engagement toy, not a stretch tool — and indistinguishable from a worksheet.
Free, low-friction, level-graded entry; a "top X% at this level" signal + a path onward	High achievement is concentrated and often unrealized; competitive instruments surface latent talent that grades and standard tests miss.	A walled garden you have to already know about: we'd be re-serving the kids who are already served, and missing the point.
Cooperative/team and solo modes are <i>out of scope</i> for the thesis	For the general K-12 population, cooperative structures beat competitive ones for both achievement and belonging — so those modes serve a different audience.	Bolting them on as v1 priorities to "broaden the market" dilutes the wedge and turns us into a crowded general app (see §4, §8). They're a later, separate bet.

THE MOVE

We're not betting that "competition is good for learning" — it isn't, on average. We're betting that **a competitive structure, engineered to trigger approach motives and avoid avoidance ones, is a strong motivational lever for confident high achievers**, and that **no consumer product is currently aimed at exactly that combination of structure and segment**. The product *is* the operationalization of the literature; the literature *is* why the product looks the way it does. (Every claim in this section is cited in docs/MathClash-Research-Competition-K12.)

4 WHAT WE BUILD — AND WHAT WE REFUSE TO BUILD

The discipline of a narrow thesis is mostly a list of "no." A feature earns its place only if it makes the thesis sentence more true for the target kid.

BUILD — V1 The thesis, made of pixels

- 1v1 duels as the primitive; matchmaking by subject + level (then a skill tier).
- A genuine difficulty ladder per subject (start: Pre-Algebra → Algebra 1 → Geometry), points scaling with difficulty, worked solutions shown after every question.
- Anonymous randomized handles for all players; report/block; profanity-safe handle generation.
- Divisoned leaderboards (your tier, near you) — not one global list; reward improvement, accuracy, streaks, not only wins.
- A "you're in the top X% at this level" signal plus an explicit **path onward** (math circles, AMC/MATHCOUNTS, acceleration resources).
- Free and low-friction to enter; consumer; no account-for-school friction in v1.

REFUSE — ON THESIS GROUNDS

- Grades, stakes, or anything where a permanent consequence rides on a loss. Never grade on rank.
- Public real-name identities or an "audience" around a match.
- One giant global leaderboard as the headline number.
- Pay-to-win, or monetization that buys competitive advantage.
- Broad K-5 gamification; "worksheet with points" content; same-difficulty grinding.
- Streak-anxiety dark patterns or engagement metrics chased at the expense of well-being.
- Claiming we "raise test scores." We claim engagement and stretch for a segment; outcomes are a pilot question, not a tagline.
- Becoming a general-purpose K-12 study app. That's a different, more crowded, worse business.

LATER — SEPARATE BETS, NOT V1

- Cooperative / team modes and a solo practice ladder (a *broader-audience* bet; out of the thesis on purpose).
- A school-side "challenge layer" districts can switch on for their high achievers (the lane the evidence says works but most districts skip).
- Higher rungs of the ladder toward contest tiers (where demand overlaps the AoPS-adjacent world) and a credentialing signal.

5 WEDGE → EXPANSION, WITHOUT BETRAYING THE THESIS

Start narrow: grades 6–9, three subjects up the early ladder, consumer, US; acquire the capable-but-unserved kid through the people who already touch them — math-club and accelerated-track teachers, talent-search-adjacent parent communities, math creators. Win retention and "are they actually stretching" first (see §6).

Expand only along axes that make the thesis more true: (1) **up the ladder** — more subjects and higher difficulty tiers, toward contest-level math, deepening the "stretch"; (2) **further along the path onward** — contest prep, acceleration, a real credentialing signal, so we become the *on-ramp to the math-talent pipeline*, not just a game; (3) **into schools, for the same kid** — a thin challenge layer for high achievers, which is the school-side version of our exact thesis.

The thing we never become: a general K-12 worksheet-gamification app, or a competition-elite walled garden. Both are off-thesis; one is crowded, the other is already served. (Phase plan, team and budget for the above: docs/MathClash-Build-Plan, §3–4.)

6 HOW WE'LL KNOW THE THESIS IS TRUE — FALSIFIABLE BETS

The thesis is a claim, so it has to be killable. Each bet below has a metric and an honest "if this fails, the thesis (as stated) is wrong" reading.

Bet 1 — It retains the target kid without school pressure.

Metric: D7 / D30 retention and matches-per-user *among the capable-but-unserved segment* (not the general signup pool). **Fails if:** they don't come back voluntarily → the "motivational structure" claim doesn't hold for the audience we can actually reach.

Bet 2 — The experience stretches, it doesn't just entertain.

Metric: share of users who climb the difficulty ladder over time; accuracy at higher tiers improving with play. **Fails if:** users plateau low and just churn matches → we're an engagement toy, not a stretch tool, and the "real difficulty ladder" clause is decorative.

Bet 3 — We reach kids the system doesn't serve.

Metric: share of active users with no gifted track / no math team / from non-elite schools (light self-report). **Fails if:** we're mostly serving kids already in AoPS/CTY → we're redundant; the "whose schools never would" clause isn't being delivered.

Bet 4 — The well-being guardrails actually hold in execution.

Metric: self-reported pressure/anxiety, rage-quit and abandon-mid-match rates, opt-out rates; qualitative parent feedback. **Fails if:** the target kid reports it makes math more stressful → "low-stakes" failed in practice and the thesis's safety condition is unmet.

Bet 5 — A near-peer market actually clears.

Metric: matchmaking wait times and the share of matches that are well-matched (not bot-padded) at each subject × level × tier. **Fails if:** liquidity never builds for a niche segment → "near-peer" is unbuildable at our scale and the model needs rethinking (async/ghost duels, narrower launch).

Anti-metrics we will not chase: DAU at all costs; time-on-app for its own sake; streak mechanics that work by inducing anxiety; "wins" inflation. If those go up while Bets 2–4 go sideways, we're winning the wrong game.

7 RISKS TO THE THESIS

RISK	MITIGATION
The segment is real but too diffuse to reach cheaply — no obvious channel to "capable-but-unserved" kids.	Test channels in Phase 0 on a small budget: math-club / accelerated-track teachers, talent-search-adjacent parent communities, math creators. If CAC won't work, the consumer wedge is wrong even if the thesis is right.
"Near-peer" needs matchmaking density a niche can't supply.	Launch deliberately narrow (one subject, two grade bands); high-quality, clearly-labeled practice bots; async "ghost" duels against recorded runs as a liquidity bridge.
Parents/schools hear "competition" and read "anxiety" — and won't allow it.	Lead with the design discipline (low-stakes, anonymous, no grades, off-ramps, parental controls) and publish the research brief; make the safety story the pitch, not a footnote.
Growth pressure drags us toward the crowded general-K12-gamification lane.	This document is the filter. Every roadmap item gets checked against the thesis sentence; "broaden the market" is not a sufficient reason to ship something off-thesis.
The evidence base is supportive but indirect — no study tested a product like this.	Hold the thesis as a falsifiable bet, not a settled fact; Bets 1–5 measure it directly in a real pilot (Build Plan, Phase 0 exit criteria). If the pilot disconfirms, the thesis changes — or the product doesn't ship.

8 THE ANTI-THESIS — WHAT WE ARE EXPLICITLY NOT CLAIMING

- Not claiming competition makes students better at math *on average* — the meta-analytic evidence says it doesn't.
- Not a replacement for instruction, a curriculum, or tutoring — we're a motivational layer, not a teacher.
- Not for the general K-12 population — for the median or struggling student, cooperative and individualized structures win; that's a different product.
- Not optimizing for engagement at the expense of student well-being — the guardrails are load-bearing, not PR.
- Not competing with AoPS/CTY for the already-identified elite — different segment, different content depth; we may become their on-ramp, not their rival.

IF YOU REMEMBER ONE THING

MathClash is a bet that there's a specific kid the education system structurally under-serves, and a specific, evidence-backed motivational structure that fits her — and that putting the two together is a product nobody else is building. Keep the thesis sentence narrow, instrument the bets, and let the pilot tell us whether to scale it, pivot it, or shelve it.