MathClash — Productization Build Plan

From front-end demo to a shippable 1v1 math-duel product for grades 6–12

Working document · drafted May 2026 · owner: Product · status: pre-build scoping

BOTTOM LINE UP FRONT

The current build is a polished **front-end demo** — the duel UX is real, but the opponent, scoring, content store and identity are all client-side stubs. That's roughly **3–5% of a real product**. A **thin but real MVP** (one subject done well, server-authoritative multiplayer, real accounts, basic trust & safety) is **~3 months** with a 2-person core team. A **credible v1** you could fundraise or run paid pilots on is **~6–9 months** with 5–6 people and a fully-loaded launch cost of roughly **$0.6–1.2M**. The dominant cost driver is not engineering — it is **content creation/licensing** and **trust-&-safety / minor-user compliance**, both of which are ongoing, not one-time.

1 THE STRATEGIC FORK — PICK THE CUSTOMER FIRST

This decision moves the estimate by 2–3×; you cannot cheaply build both for v1.

PATH	WHAT IT DEMANDS	HARD PART	REVENUE SHAPE
Consumer (B2C / freemium)	Retention loops, daily streaks, social/viral hooks, high content volume, light-touch parental-consent flow (COPPA). App-store presence.	Distribution & retention — you live or die on growth marketing and the content treadmill.	Subscriptions / IAP. Faster to launch, noisier, lower per-user value.
Schools (B2B2C)	Rostering (Clever / ClassLink / Google Classroom), teacher & admin dashboards, standards alignment (Common Core / state), per-district DPAs, SOC 2, accessibility (WCAG), procurement support.	Sales cycle & compliance — 6–12 month district sales, heavy build before first dollar.	Seat / site licenses. Slower, but sticky and higher LTV.

Recommendation: launch **consumer-first** to maximize speed of learning and keep the build narrow; design the data model and content metadata so a school/dashboard layer can be added later without a rewrite. Switch the priority order only if you have a warm path into districts.

2 GAP ANALYSIS — WHAT'S ACTUALLY MISSING

DONE IN DEMO · PARTIAL / STUBBED · NOT STARTED · NET-NEW WORKSTREAM

WORKSTREAM	STATUS	EFFORT TO V1	NOTES / WHY IT MATTERS
Duel UX & game loop	DONE	~done	1v1 screen, two modes, difficulty ladder, 60s timer, leaderboard, anonymous handles — all built. Needs polish, not invention.
Real-time multiplayer	STUBBED	4–8 wks	Opponent is a `setTimeout` today. Need a matchmaking service (queue by subject + level + skill + latency), **server-authoritative** game state, reconnection & "opponent left" handling. Core of the product; first thing that breaks at scale.
Content library & pipeline	60 QS	2–6 mo +ongoing	60 Pre-Algebra items today; a real product needs **thousands** across 10+ levels, tagged by topic/standard/difficulty, with worked solutions and misconception-mapped distractors. Requires curriculum hires or licensing, a review workflow, and an authoring CMS. **This is the moat and the cost center.**
Accounts / auth / minors	NONE	3–6 wks + legal	localStorage today. Grades 6–12 ⇒ many users **under 13 (COPPA)**: verifiable parental consent, data-retention limits, restricted analytics/ads. School path adds **FERPA** + state laws (SOPIPA etc.) + signed DPAs. Needs real counsel, not a template.
Trust & safety	HANDLES	2–4 wks + ops	Randomized handles in place. Add report/block, profanity-safe handle generation, rate-limiting, abuse/grief detection, **no free-text chat in v1** (or heavily filtered), and a human to action reports. Kids' multiplayer attracts grief — underinvesting here is a headline risk.
Anti-cheat / integrity	NONE	2–4 wks	Answer key is literally in `questions.js` now. Move answers/scoring server-side, keep pools large enough to resist memorization, detect impossible streaks, sanity-check the leaderboard. Scales with any prize/status attached to rank.
Placement & profiles	NONE	2–3 wks	Onboarding placement quiz (so a 7th grader isn't matched into Calculus), persistent progress/history/streaks backed by a DB, friends & private matches.
Mobile	RESPONSIVE WEB	0 / 6–10 wks	Responsive web covers Chromebooks (the school workhorse). Phones want a polished PWA at minimum, likely React Native / Flutter — plus kids'-app store review.
Infra / analytics / payments	NONE	3–6 wks + ongoing	Scale for spiky concurrency (school bell schedules = thundering herds), observability, a **minor-compliant** analytics stack, Stripe / Apple / Google billing or district invoicing, CI/CD, on-call.
Teacher / parent dashboard	B2B ONLY	6–10 wks	Only if school path: class views, assignments, progress reports, standards coverage. Skip for consumer v1.
Live-ops & balancing	FOREVER	ongoing	ELO tuning, difficulty-curve tuning, "the leaderboard feels unfair" tickets, seasonal events, A/B tests. Never finishes; it's the job.

3 PHASED ROADMAP

Three gates. Don't start a phase until the prior phase's exit criteria are met — especially: don't build native apps or dashboards before retention data says the loop works.

Phase 0 — Thin Real MVP ~3 MONTHS

Team: 1 senior full-stack + 0.5 curriculum + PM (you)

GOAL	Replace every stub with the minimum real version; put ~100 real students on it and watch what happens.
IN SCOPE	Server-authoritative real-time multiplayer + matchmaking (managed: Colyseus / Supabase Realtime / PlayFab); real accounts (incl. a basic COPPA parental-consent flow); move answer key & scoring server-side; one subject built deep (~400–600 calibrated Pre-Algebra items + worked solutions); placement quiz; persistent stats & leaderboard in a DB; report/block + safe handles; basic analytics; web only (keep it responsive).
OUT OF SCOPE	Native apps, teacher dashboards, payments, multiple subjects, friends/social, seasonal events.

EXIT CRITERIA → D1/D7 retention and matches-per-user clear an internal bar; matchmaking wait times acceptable; zero unhandled safety incidents in pilot; content review process working.

Phase 1 — Credible v1 +3–6 MONTHS

Team: 2–3 eng + 1 designer + 1–2 curriculum + PM

GOAL	A product you'd put in front of investors or run a paid pilot with — broad enough to be "real," polished enough to retain.
IN SCOPE	3–5 subjects live (Pre-Algebra → Algebra 1 → Geometry → …, content treadmill running); ELO/skill-based matchmaking with anti-cheat & integrity checks; profiles, friends, private matches, rematch; re-engagement (push / email, compliant); polished PWA for phones; full analytics + experimentation; monetization (subscription / IAP); accessibility pass; legal review complete (COPPA, ToS, privacy).
OUT OF SCOPE	District rostering & dashboards, SOC 2, native app-store apps (unless data demands it).

EXIT CRITERIA → retention & monetization metrics support a growth or fundraise decision; content backlog is being produced faster than it's consumed; CAC/LTV directionally works on at least one channel.

Phase 2 — School-ready / Scale +6–9 MONTHS

Team: 5–8 eng + content team of 3+ + design + sales/CS

GOAL	Sellable to districts and able to take real concurrent load — competing with IXL / DeltaMath / Prodigy on their turf.
IN SCOPE	All 10 levels with depth; rostering (Clever / ClassLink / Google Classroom); teacher & admin dashboards (assignments, progress, standards coverage); explicit standards alignment; SOC 2 Type II; per-district DPAs & security review support; native iOS/Android; scaled infra & on-call; support org; live-ops cadence (events, seasons).

EXIT CRITERIA → first district contracts signed; concurrency SLOs holding through peak (bell-schedule) load.

4 TEAM & BUDGET

Headcount ramp

- **Phase 0:** 1 senior full-stack eng, ~0.5 curriculum/content, PM (you), fractional designer + legal counsel.
- **Phase 1:** +1–2 eng, +1 product designer, +1 curriculum lead, contract content writers.
- **Phase 2:** eng team of 5–8 (incl. infra/SRE + data), content team of 3+, sales + customer success, support, compliance/security lead.

What the money actually buys

Engineering is **not** the dominant line. Recurring spend is led by **content creation/licensing**, then **trust-&-safety + moderation ops**, then **infra**. Treat content as a standing team, not a project sprint.

Rough fully-loaded cost (US, blended)

~$0.15–0.35M PHASE 0 → THIN MVP LAUNCH	**~$0.6–1.2M** THROUGH PHASE 1 → CREDIBLE V1 LAUNCH
$2M+ PHASE 2 FIRST-YEAR RUN (SCHOOL-READY)	**3–5 / 5–6 / 8–14** FTE-EQUIV: PHASE 0 / 1 / 2

Ranges, not quotes — swing with US vs. blended/offshore eng, how much content is built vs. licensed, and whether the school path is pulled forward.

5 TOP RISKS & MITIGATIONS

RISK	WHY IT BITES	MITIGATION
Content is the real product — and it's slow & expensive	Most "ed-tech platforms" are content companies with an app attached. Thin or low-quality banks kill retention and invite memorization-cheating.	Start the curriculum pipeline on day one, in parallel with eng. Hire a curriculum lead early. Consider licensing a starter bank to bridge. Build the authoring CMS before the bank gets large.
Minor-user compliance (COPPA / FERPA / state laws)	Not optional, not a checkbox — it shapes signup, logging, analytics, ads, retention, and (for schools) every contract. Getting it wrong is existential.	Engage specialist ed-tech counsel before building auth. Design "privacy-by-default" data model. Keep analytics minor-safe from v1. Budget real legal hours each phase.
Trust & safety incident	Kids' multiplayer attracts grief, harassment, and worse; one bad incident becomes a news story and a trust collapse.	No free-text chat in v1. Safe handle generation, report/block from launch, rate-limiting + abuse detection, a real human on report queues, and an incident playbook.
Cheating / leaderboard integrity	Client currently holds the answer key; any status or prize on rank invites scripted play and erodes the competitive core.	Server-authoritative answers & scoring; large rotating pools; streak/accuracy anomaly detection; periodic leaderboard audits.
Cold-start / matchmaking density	1v1 only works if there's someone to match with at your subject + level + skill; thin liquidity ⇒ long waits or bad matches.	Launch narrow (one subject, one or two grade bands). High-quality bots clearly labeled as practice while liquidity builds. Async/"ghost" duels against recorded runs as a fallback.
Spiky concurrency (bell schedules)	Usage isn't smooth — it spikes at class periods; naive infra falls over exactly when schools are watching.	Load-test against bell-schedule patterns; autoscaling + queueing; graceful degradation; SLOs tracked from Phase 1.
Competitive moat	IXL, DeltaMath, Prodigy, Khan Academy own mindshare; "math questions in an app" alone isn't defensible.	Lean into the wedge the demo already has — **live, social, head-to-head competition** — plus content depth and curated difficulty. Be the place math feels like a game, not a worksheet.

6 NEXT 90 DAYS — CONCRETE ACTIONS

WHEN	ACTION
Weeks 1–2	Lock the customer decision (consumer vs. school). Engage ed-tech counsel for a COPPA/FERPA scoping memo. Spec the data model so a dashboard layer can bolt on later. Pick the real-time stack (build vs. managed) and stand up a spike.
Weeks 2–4	Hire / contract the senior full-stack eng and a part-time curriculum lead. Stand up auth + DB + server-authoritative game state. Move the answer key and scoring off the client.
Weeks 3–8	Build Pre-Algebra to depth (~400–600 calibrated items + worked solutions) with a lightweight authoring workflow. Ship matchmaking, placement quiz, persistent stats/leaderboard, report/block.
Weeks 8–12	Closed pilot with ~100 students (a school club, a tutoring network, or a creator audience). Instrument retention, match wait times, accuracy curves, safety reports. Review against Phase 0 exit criteria → decide go / iterate / stop.

IN ONE LINE

Treat the demo as a validated **UX prototype**, not a head start on the backend: ~3 months & two people to a real MVP, ~6–9 months & ~$1M to a credible v1, and the long pole isn't code — it's **content depth** and **doing right by a userbase of minors**.

7 BUILD IT YOURSELF VS. BUILD IT WITH AN ESTABLISHED PLATFORM

Partnering with an existing teaching platform doesn't make the *product* smaller — it makes **your share of the work** smaller, by handing you the three things that dominate the greenfield budget: a content library, an engaged audience (no cold-start), and a compliant home for minors. The price is equity and control. Four shapes the deal can take:

PATH	WHAT YOU GET	WHAT YOU GIVE UP	TO V1
Greenfield **(your company, build it all)**	Full control of product, brand, roadmap, and upside. The standalone-consumer outcome if it works.	You bear every cost & risk — content team, cold-start, compliance, infra, marketing.	~6–9 mo ~$0.6–1.2M
Inside AoPS **(their product; you build / contract)**	Their problem bank (Alcumus-style), their community, their brand, their COPPA infra — and they've built real-time HTH math before (*For the Win!*). You're shipping to a real audience on day one.	You own ~nothing. It's their roadmap, their pedagogy, their brand rules. Realistically a hire / services / acqui-hire outcome, not a co-owned company.	~2–4 mo integration
Partner with DeltaMath / a curriculum publisher **(your company, their content + distribution)**	Kills the content build and the cold-start; publishers also bring district relationships. You keep the company.	Rev-share, co-branding constraints, roadmap influenced by the partner, and they can de-prioritize / churn you. Publishers move slowly.	~3–5 mo
Embed in a Khan-style non-profit	Enormous general-audience reach and a mission halo; their accounts & trust.	No equity / acquisition story; you're a feature on someone else's mission; deliberate pace.	~3–5 mo

What a content + audience partner actually removes

WORKSTREAM (FROM THE GAP ANALYSIS)	GREENFIELD	WITH A CONTENT + AUDIENCE PARTNER
Content library & pipeline	2–6 mo + a standing curriculum team	**≈ 0** — license their categorized bank; reframe "subject + difficulty" onto their taxonomy
Cold-start / matchmaking liquidity	Slow and risky — 1v1 needs density at every subject + level + skill tier	**Solved** — drop into an existing engaged community
Accounts / auth / minor compliance (COPPA & co.)	3–6 wks + specialist counsel + privacy-by-default data model	**Mostly inherited** — they already operate at scale with minors
Real-time multiplayer	4–8 wks to build server-authoritative matchmaking & game state	**Reduced** — reuse their patterns / infra (AoPS has shipped this with *For the Win!*)
Brand & trust	Earned over years	**Borrowed on day one**
Net effect	**~6–9 mo · ~$0.6–1.2M** greenfield to a credible v1	**~2–4 mo of integration work** — but it's no longer "your" product

8 THE AOPS CASE, SPECIFICALLY — AND BETTER-FIT PARTNERS

THE AOPS REALITY CHECK

AoPS is a strong illustration but a **tricky** partner for *this specific product*: (1) they **already have a real-time head-to-head math game** (*For the Win!*) plus Alcumus, the Community forums, the textbooks, and Beast Academy — so you'd be pitching "FTW 2.0," an internal-product replacement, not a new idea; (2) they're a **build-it-ourselves, brand-protective** shop, and "gamified battles" sits somewhat *against* their "real math, not worksheet gamification" positioning; (3) their audience is the **competition-track top ~5–10%**, not the general grades 6–12 (pre-algebra → calculus, curriculum-aligned) lane MathClash targets — and that mainstream lane (Khan / DeltaMath / IXL / Prodigy) is precisely the one AoPS has chosen *not* to serve. The content model would also shift from a curriculum ladder to a **contest-tier ladder** (AMC 8 → AMC 10 → AIME …), which actually suits that audience better — but it's a different product than the one specced.

So with AoPS specifically: don't pitch a 50/50 partnership — pitch **building their next-gen competitive arena** as a contractor / hire / small acqui-hire, using the demo as the working prototype; *or* build MathClash independently for the mainstream lane AoPS avoids and treat AoPS / IXL / a publisher as a later **acquirer**, not a Day-1 partner.

If the goal is "an established platform + my competitive layer," better-fit partners than AoPS

PLATFORM	FIT / WHY	CATCH
DeltaMath	Huge in US high-school math classrooms, teacher-driven, small team, **no consumer / social / competitive layer at all** — "challenge a classmate" is a natural extension they haven't built. Most realistic "yes."	Small team; B2B-classroom flavored; a consumer motion would be new for them.
Khan Academy	Massive general grades 6–12 audience, mission-aligned, **no real competitive multiplayer**; their accounts & trust.	Non-profit — reach, not equity; deliberate pace.
Curriculum publisher (McGraw Hill, Savvas/Pearson, HMH)	They have the content **and** the district relationships, and want "engagement" layers badly.	Slow, bureaucratic; co-branding strings; long contracts.
IXL	Ideal audience; already heavily gamified (SmartScore, awards); B2B school-sales machine.	Most likely to **just build it themselves** rather than partner.
Prodigy	—	Already does gamified math "battles" for K–8. A **competitor**, not a partner.

IN ONE LINE

A platform deal can take v1 from **~6–9 months / ~$1M greenfield** to **~2–4 months of integration** — but you trade nearly all the equity for nearly all the de-risk, and the partner has to actually want it. **AoPS** is more plausible as a hire / acqui-hire ("FTW 2.0") or a future acquirer than as a Day-1 partner; **DeltaMath or a publisher** is the more realistic "yes" if you want to keep a company while shedding the content and cold-start risk.

9 APPENDIX — LEAN MVP VARIANT ("PHASE 0-LITE", ~2–6 WEEKS)

Drop the three pillars that dominate the Phase 0 estimate — real-time multiplayer, the "for school children" positioning, and the curriculum-team content build — and replace them with what the demo already has plus AI-drafted, founder-reviewed content. The timeline collapses from ~3 months (Phase 0) to **~2–6 weeks**, with essentially **no backend**. Caveat: this is *not* a test of the competition thesis (the human opponent and the leaderboard are exactly the "near-peer 1v1 competition" pillars — see Product Thesis §3 & §8); it's a cheap test of the **stretch-ladder + percentile** idea, and it converges toward an Alcumus-style solo product. Use it as a learning vehicle, then layer Phase 0 / Phase 1 on top if it shows life.

ITEM	PHASE 0 (THIN REAL MVP)	LEAN MVP	CHANGE
Positioning	"for grades 6–12" — COPPA in scope	—	General-audience practice app, **not** a children's product → no verifiable-parental-consent requirement, no specialist counsel.
Front-end / UX	polish	~3–7 d	Demo already has home / vs-bot match / results / stats. Drop grade selection; swap leaderboard → "≈ top X%" panel; strip school copy; polish.
Opponent	4–8 wks (multiplayer + matchmaking)	~0–2 d	Keep the demo's 70–100% accuracy-variance bot as a pacing partner; retune only. The hardest, most scaling-prone piece of Phase 0 — removed.
Accounts / minor compliance	3–6 wks + specialist counsel	~0–3 d	No accounts; progress in `localStorage`. Boilerplate ToS/privacy + "not directed to children under 13" — a template, not a project.
Leaderboard / cold-start	(the liquidity risk)	~1–3 d	None. Show an estimated percentile from an anchored difficulty→accuracy curve, clearly labeled an estimate.
Content (~200 Qs × N categories, AI-gen + founder review)	2–6 mo + curriculum team	**bottleneck**	Generation/formatting fast; **founder review is the real project** — ≈1–2 min/Q (more for hard items) ⇒ ~30–60 focused hrs for the full 10×200; expect a non-trivial AI-gen error rate (wrong keys, ambiguous distractors, mis-tagged difficulty, near-dupes, notation). Trim the first ship to 3–4 categories deep (~600–800 Qs).
Backend / infra	3–6 wks	~0	No MP, no leaderboard, no accounts ⇒ ship static like the demo; client-side answer key is fine (no stakes to cheat for).
Analytics	—	~0.5 d	A privacy-friendly tracker (Plausible / Umami) so you can tell if anyone uses it.

9 APPENDIX — LEAN MVP VARIANT (CONTINUED): TIMELINES, WHAT IT BUYS, RISKS

~2–3 wks
TRIMMED FIRST SHIP · 3–4 CATEGORIES, ~600–800 QS · ~1.5–2 WKS OF IT CONTENT REVIEW (A FEW WEEKENDS IF FOUNDER-BUILT WITH AI)

~4–6 wks
FULL 10 CATEGORIES × ~200 QS · ~3–4 WKS OF IT CONTENT REVIEW

+1–3 wks
OPTIONAL THIN BACKEND LATER · MAGIC-LINK ACCOUNTS, SERVER-SIDE CONTENT, ANALYTICS DASHBOARD

⇒ roughly **5–10× faster than Phase 0** — every item cut was a top cost driver, not a peripheral one. **Validates:** appetite for a progressively-harder solo ladder + whether a "top X%" signal motivates + content quality + the difficulty curve. **Does not validate:** the competition thesis — the bot is a liquidity bridge, not a substitute. **Sequencing:** lean MVP → instrument retention + ladder-climb + percentile believability → if it shows life, add real 1v1 + a leaderboard (= the spec'd Phase 0) → Phase 1. If it doesn't retain, you learned that in weeks, not months.

RISKS SPECIFIC TO THE LEAN VARIANT

(1) AI-generated math at volume *will* contain errors — the founder review is the actual deliverable; consider a second reviewer or a sampling-based QA gate before each batch goes live. **(2)** A percentile shown without real user data can be embarrassingly wrong — anchor it to a defensible per-difficulty accuracy assumption, label it an estimate, and swap in real percentiles once there's traffic. **(3)** Don't conflate "the lean MVP retained" with "the competition thesis is validated" — keep the two questions, and the two roadmaps, separate.