Peer competition in K–12 education — and what it does for high-achieving students

A research brief synthesizing meta-analytic and quasi-experimental evidence

Prepared for MathClash · drafted May 2026 · all quantitative claims cited to primary sources (see References, p. 7) · this is a focused synthesis, not a systematic review

Abstract. Does competing against peers help students learn — and does it help the strongest students in particular? The best evidence says: **competition has essentially no average effect on performance by itself** [1]; its sign depends on which motive it activates — a desire to *outperform* others (which tends to raise performance) versus a fear of *looking incompetent* (which tends to lower it) [1,2,3]. That conditionality is exactly why competition tends to skew **positive for confident, high-achieving students** and negative for the anxious and the struggling. Quasi-experiments back this up: grouping high achievers into selective, ranked classrooms raised their scores by ~0.5 SD [4]; a student's ordinal rank in class has lasting effects on later achievement and confidence independent of ability [5]; giving students relative-performance feedback modestly raised their grades [6]. Competition data (e.g., the American Mathematics Competitions) also *reveals* how concentrated and how often unrealized high math talent is [7,8], and longitudinal work on early-identified talent shows extraordinary downstream STEM and creative output when ability is identified and then stretched [9,10]. But the countervailing evidence is real: across *all* students, cooperative structures beat competitive ones for both achievement and peer relationships [11]; classroom leaderboards can erode intrinsic motivation over a semester [12]; and the same pressure that lifts scores is associated with lower life satisfaction, especially among top performers [13]. The practical conclusion: well-designed, **near-peer, low-stakes, repeated** competition — with a mastery anchor and losing framed against an absolute standard — is one of the more reliable motivational levers for high achievers [14,3]; broad-N, high-stakes, public, win-or-fail competition is the version that backfires.

0 THE HEADLINE, STATED PLAINLY

- **"Does competition help kids learn?" is the wrong question.** Meta-analysis of decades of experiments finds **no noteworthy direct relation** between competition and performance [1]. Competition is a context, not an intervention; what matters is the motive it triggers.
- **It splits people.** Competition reliably pushes some toward *performance-approach* goals ("I want to do better than others"), which predict better performance, and others toward *performance-avoidance* goals ("I don't want to be the one who looks bad"), which predict worse performance [1,2]. The split tracks confidence and prior success — so the net effect is, on average, **more favorable for high achievers**.
- **For high achievers specifically, the causal evidence is encouraging.** Selective grouping/tracking of high achievers [4], a high ordinal rank among peers [5], and relative-performance feedback [6] all produce measurable, often persistent gains.
- **Design is decisive.** Small numbers of competitors raise competitive drive more than large ones (the "N-effect") [14]; framing a loss against an absolute standard with positive feedback keeps "losers" motivated [3]; raw leaderboards-and-badges, by contrast, measurably demotivated students over a term [12]. Same ingredient, opposite outcomes, depending on the recipe.
- **It is not a free lunch.** Cooperative learning still out-performs competitive learning for the typical student [11], and competitive pressure carries a well-being cost concentrated among the students who already feel it most [13]. A responsible product offers competition as *one* mode, not the only one.

EVIDENCE-QUALITY NOTE

Strength of inference varies. META-ANALYSIS pooled experiments/studies (refs 1, 2, 11). QUASI-EXPERIMENTAL / CAUSAL regression-discontinuity, natural experiments, or controlled experiments (refs 1, 3, 4, 5, 6, 12). DESCRIPTIVE / CORRELATIONAL large-scale observational patterns (refs 7, 8, 9, 10, 13). Where a claim rests on correlational data, the text says so.

1 COMPETITION'S EFFECT ON PERFORMANCE IS CONDITIONAL, NOT CATEGORICAL

The meta-analytic baseline: roughly zero — because two effects cancel

Murayama & Elliot's *Psychological Bulletin* meta-analysis is the anchor result. Pooling the existing experimental literature, they found **no meaningful overall relationship between competition and performance**. But that null is not "competition does nothing" — it is two opposing pathways netting out: competition tends to elicit *performance-approach* goals, which **facilitate** performance, and simultaneously elicit *performance-avoidance* goals, which **undermine** it. Strip away the averaging and competition is doing a lot — just in opposite directions for different people and situations.[1]

Which "performance goal" you measure changes the sign

Achievement-goal research refines this. Hulleman and colleagues' meta-analysis showed that **how** a performance-approach goal is operationalized matters: **normative** performance-approach goals — wanting to *outperform others*, which is precisely what a duel or a leaderboard makes salient — were **positively** related to performance, whereas **appearance** performance-approach goals — wanting to *look smart / avoid looking dumb* — were **negatively** related.[2] A later meta-analysis (Senko & Dawson) reached the same conclusion: performance-approach goal effects "depend on how they are defined," with the competitive/normative version the benign one.[3a] Implication for product design: keep the framing on "*beat your match / climb the ladder*" (normative) and away from "*don't embarrass yourself in front of others*" (appearance) — e.g., via anonymity and low stakes.

Winning, losing, and whether losers stay in the game

Vansteenkiste & Deci tested the obvious worry — that competition demotivates whoever loses. Winners were indeed more intrinsically motivated than losers. **But losers who were given a clear, absolute performance standard and positive feedback for meeting it stayed about as intrinsically motivated as winners**; losers given a tangible "performance-contingent reward" for the same standard did *worse* on a behavioral measure of motivation than those simply told they'd done well.[3] Translation: a competitive system survives the existence of losers *if* "you lost the match" is not the same message as "you failed" — pair every duel with an absolute yardstick ("you nailed 7/10, and the level-8 question") and concrete next steps, not just a W/L.

FINDING	EVIDENCE	SO WHAT FOR A COMPETITIVE LEARNING PRODUCT
No average effect of competition on performance; net of opposing approach/avoidance pathways [1]	META-ANALYSIS + EXPERIMENTS	Don't claim "competition boosts learning." Claim "competition, designed to trigger approach (not avoidance) motives, boosts engagement and effort for the students it suits — disproportionately the confident/high-achieving."
Normative ("outperform others") performance-approach goals → better performance; appearance ("look able") → worse [2,3a]	META-ANALYSIS	Frame around rank and head-to-head outcomes; suppress public identity / "audience" cues that activate appearance concerns (anonymous handles do exactly this).
Losers stay motivated if given an absolute standard + positive feedback; tangible "you-lost-so-no-prize" rewards hurt [3]	EXPERIMENT	Always show an absolute score and a worked solution alongside the match result; reward improvement/streaks/accuracy, not only wins.

2 WHERE COMPETITION RELIABLY HELPS: HIGH ACHIEVERS, IN NEAR-PEER SETTINGS

If the average effect is a wash, the interesting question is *for whom* and *under what conditions* it's positive. The causal evidence converges on: **high-achieving students, compared against near-peers, with rank made visible**.

Grouping high achievers together raises their scores

Card & Giuliano studied a large urban district that creates a separate "gifted / high-achiever" classroom in any school with at least one gifted fourth-grader — most seats filled not by formally "gifted" students but by the **next-highest scorers, ranked by the prior year's test scores** (an explicitly competitive entry rule). Using a rank-based regression-discontinuity design, they found the high achievers who got into these rooms gained about **0.5 standard deviations in fourth-grade reading and math**, with gains persisting through sixth grade; the effects were concentrated among Black and Hispanic participants, and there were **no negative spillovers** on the students left behind.[4] A competitive selection rule, plus a peer environment of similarly high achievers, produced a large, durable, equity-positive effect.

Your ordinal rank among peers shapes your trajectory — beyond your ability

Murphy & Weinhardt, using the universe of English primary-school pupils, isolated the effect of a student's **ordinal rank within their class** holding actual achievement constant (the same score can be 3rd in one class, 15th in another, by chance of cohort). Being ranked higher produced **large lasting effects on later test scores, confidence, and subject choice** in secondary school — even though the new secondary teachers and peers had no knowledge of the old ranking. The mechanism is non-cognitive (a confidence/identity effect, not extra learning), and the effect is **especially large for boys**, contributing to the gender gap in who later chooses STEM.[5] Being "the strong one" in a peer comparison is itself a treatment — which is the experience a competitive product can deliberately manufacture for students who'd never be top of their physical classroom.

Telling students where they stand nudges performance up

Azmat & Iriberri exploited a natural experiment in which a high school began printing each student's grade **alongside the class average**. Adding that relative-performance information **raised students' grades**, though the boost attenuated over time.[6] Simple social-comparison feedback — the core of any leaderboard — has a real, if modest and possibly fading, positive effect in a real school setting.

"Number of rivals" matters: small-N competition is more motivating

Garcia & Tor's "N-effect": as the number of competitors grows, **competitive motivation falls**. Across studies, average test scores were lower at larger test-taking venues; people told to finish in the top 20% raced faster against a field of 10 than a field of 100; the effect runs through social comparison and is strongest in people high in social-comparison orientation.[14] A 1-versus-1 duel sits at the maximally motivating end of this curve — and small leagues/divisions beat one giant global leaderboard.

Talent development needs the psychosocial muscle that competition builds

The gifted-education research community has converged on the view that translating raw ability into eminent achievement depends heavily on **psychosocial skills** — motivation, a willingness to take intellectual risks, and the ability to **compete and to cope with setbacks and criticism** — and that these are malleable and need to be cultivated through opportunities for productive challenge (Subotnik, Olszewski-Kubilius & Worrell, in *Psychological Science in the Public Interest*).[15] Competition, used deliberately and supportively, is one of the standard vehicles for that.

3 COMPETITION AS A DISCOVERY — AND STRETCHING — MECHANISM FOR TALENT

Beyond motivating individuals, competitions function as **instruments**: they surface high-ability students who standard assessments miss, and (when paired with appropriate challenge) they help convert that ability into accomplishment.

Math competitions reveal how concentrated — and how unrealized — top talent is

Ellison & Swanson used the Mathematical Association of America's American Mathematics Competitions (AMC) — a precalculus contest taken by >100,000 U.S. students at ~3,000 high schools — because **ordinary data top out before the achievement levels that matter**. They documented "large differences in the frequency with which students from seemingly similar schools reach high achievement levels," including a **"thick tail" of schools that produce many more high achievers than typical** — a pattern not explained by observable student characteristics and persistent over time, which they read as evidence of **school effectiveness** (some environments systematically turn high ability into high achievement; most don't).[7] A companion paper showed the gender gap among the highest math achievers is far larger than standard data reveal, and that the highest-achieving girls are clustered in a small set of elite schools — implying **most girls with the latent ability to reach top math levels are not getting there**.[8] Competitive instruments make the latent-talent problem visible; a low-friction, opt-in competitive platform is one way to widen who gets the chance to find out they're in that tail.

Identify early, then stretch — and the downstream output is extraordinary

The Study of Mathematically Precocious Youth (SMPY) — a planned 50-year longitudinal study — tracks people identified in early adolescence (around age 12–13) via **above-level testing**, an inherently competitive screening. Its most selective cohort, the "top 1 in 10,000" (SAT-Math ≥ 700 and/or SAT-Verbal ≥ 630 as seventh-graders; N ≈ 320), pursued doctorates at **more than 50× base-rate expectations** in a 10-year follow-up, several already having produced notable scientific or literary work by their early 20s.[9] In later follow-ups this cohort had, by their late 30s, accumulated **392 refereed STEM publications, 820 software/patent contributions, 128 creative written works, >1,000 fine-arts achievements, and >$25M in grants**, with **11.3% holding academic tenure** (7.5% at research-intensive universities); a four-decade follow-up reported **~44% holding a doctorate** (Ph.D./M.D./J.D.) versus ~2% of the general population (Kell, Lubinski & Benbow; Vanderbilt).[10] SMPY's recurring practical finding is that the precocious do best when **allowed to advance at their own pace and given genuinely challenging work** rather than held to the standard track.[9,10] Net: competitive identification is a powerful talent radar, and the payoff comes from coupling it with stretch.

MECHANISM	EVIDENCE	RELEVANCE TO A COMPETITIVE K–12 MATH PRODUCT
Competitions surface high achievers that conventional tests/grades miss; high achievement is concentrated and often environment-dependent [7,8]	DESCRIPTIVE	A free, opt-in, level-graded competitive platform is a wide-net "talent radar" — and a way to give a student in a non-elite school the stretch experience an elite school would.
Early competitive identification + acceleration/challenge → outsized later STEM & creative accomplishment [9,10]	LONGITUDINAL	Pair the competition with a real difficulty ladder and worked solutions; the value is in the stretch, not the trophy. Don't cap strong students at "grade level."

4 THE COUNTERVAILING EVIDENCE — WHAT GOOD DESIGN HAS TO RESPECT

An honest brief has to give equal weight to what cuts the other way. None of it kills competition as a feature; all of it constrains how you build it.

For the typical student, cooperation beats competition

The largest synthesis on this question — Roseth, Johnson & Johnson in *Psychological Bulletin*, pooling **148 studies of 17,000+ early adolescents across 11 countries and eight decades** — found that **both higher achievement and more positive peer relationships were associated with cooperative goal structures rather than competitive or individualistic ones**.[11] So competition is not the default-best motivational structure for a classroom or a general K–12 population; it is a structure that suits a subset (notably the confident and high-achieving) and a particular set of conditions. A platform that *only* does 1v1 is leaving achievement and belonging on the table for many users — argue for cooperative/team modes alongside duels.

Leaderboards and badges, naively applied, demotivated students over a semester

Hanus & Fox ran a 16-week controlled comparison: one section got a gamified curriculum with a **leaderboard and badges**, an otherwise-identical section did not. Over the term, the gamified students **declined** in intrinsic motivation, satisfaction, and feelings of empowerment relative to the comparison group, and **scored lower on the final exam** — an effect statistically mediated by their reduced intrinsic motivation.[12] The lesson is not "no leaderboards"; it is that extrinsic, social-comparison-heavy mechanics can crowd out intrinsic interest if they become the point. Keep competitive mechanics **opt-in**, **low-stakes**, and **secondary to the math itself**; don't grade on rank.

The pressure that lifts scores carries a well-being cost — heaviest on top performers

OECD's PISA 2018 analysis of 15-year-olds found that, at the system level, the more **fear of failure** students reported, the higher that system's average reading scores; at the individual level, in most countries students with greater fear of failure scored higher **but reported lower life satisfaction** — and the fear-of-failure gap (girls > boys) was **wider among top-performing students**.[13] Performance-evaluative pressure and achievement often move together, but so do pressure and unhappiness, disproportionately for the students who are already most invested. A competitive product aimed partly at high achievers should treat anxiety as a first-class design constraint, not an afterthought (low stakes, no permanent "losses," easy off-ramps, no doom-scrolling the leaderboard).

> **NET READ OF THE COUNTER-EVIDENCE**
>
> Competition is a **targeted** tool, not a general one. It is most defensible when it is opt-in, near-peer, low-stakes, frequent-and-low-consequence, anonymous, anchored to an absolute standard and a learning ladder, and offered **alongside** cooperative and solo modes — i.e., when it's engineered to trigger "I want to beat my match" and not "I'm scared I'll be the loser." It is least defensible as a high-stakes, public, winner-take-all, grade-determining ranking. The same evidence base that supports the feature draws those lines around it.

5 DESIGN IMPLICATIONS — MAPPING THE EVIDENCE ONTO A COMPETITIVE K–12 MATH PRODUCT

Pulling the threads together, here is what the literature actually prescribes for a product like MathClash, and where the demo already aligns.

DESIGN CHOICE	WHY — EVIDENCE	STATUS IN THE MATHCLASH DEMO
Match on level/skill so duels are near-peer	Rank/social-comparison effects work through *relative* standing among comparable others; a high ordinal rank vs. near-peers is itself motivating [5,6], and a hopeless or trivial mismatch is not.	**Aligned** — matchmaking is by subject + level (and is designed to extend to a skill/ELO tier).
Small-N contests: 1v1 duels, small leagues	The N-effect — competitive motivation falls as the number of rivals rises; 1v1 is the maximally motivating end of the curve [14].	**Aligned** — core mode is 1v1; leaderboard should be divisioned, not one global list.
Anonymous handles; no public "audience"	Normative performance-approach goals ("outperform") help performance; *appearance* goals ("don't look dumb in front of people") hurt it [2,3a]. Anonymity also defuses harassment.	**Aligned** — randomized handles for every player, the user included.
A real difficulty ladder + worked solutions inside every match	The talent payoff is in the *stretch*, not the trophy; precocious students do best given genuinely harder work [9,10]; an absolute standard keeps "losers" motivated [3].	**Partly** — difficulty climbs 1→10 and points scale with it; worked-solution explanations are a needed addition (currently just shows the correct choice).
Reward improvement, accuracy, streaks — not only wins	Losers given positive feedback against an absolute standard stay motivated; tangible "you-lost-no-reward" mechanics backfire [3].	**Partly** — stats track accuracy/streak/best score; result screen should lead with the absolute score + what you learned, then the W/L.
Opt-in, low-stakes, frequent; nothing permanent rides on a loss; never grade on rank	Leaderboards/badges as the point demotivate over time [12]; performance-evaluative pressure tracks lower well-being, worst for top performers [13].	**Aligned** — consumer, no grades; keep stakes light and add easy off-ramps; avoid streak-anxiety dark patterns.
Offer cooperative / team and solo modes too	For the general K–12 population, cooperative structures beat competitive ones for achievement *and* belonging [11]; competition suits a subset.	**Gap** — both current modes are head-to-head; a co-op/team mode (and a solo practice ladder) would broaden the audience the evidence says competition alone won't serve.
Treat it as a wide "talent radar": free, easy to enter, level-graded	High achievement is concentrated and often unrealized; competitive instruments reveal latent talent that grades/tests miss [7,8].	**Aligned in spirit** — low-friction entry; a "you're in the top X% at this level" signal (and a path onward) would operationalize it.

ONE-LINE TAKEAWAY FOR THE PRODUCT THESIS

The research supports a specific, narrow claim — not "competition makes kids better at math," but: **"near-peer, low-stakes, anonymous, repeated 1v1 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 — and a way to give a stretch-and-rank experience to capable kids whose schools never would."** That is a defensible wedge. The broader market still wants cooperative and solo modes, and the well-being caveats are load-bearing, not decorative.

6 REFERENCES

All links retrieved May 2026. Where the publisher version is paywalled, an open-access copy or working paper is given. Page/volume details are reproduced from the cited source records; the substantive claims in this brief are drawn from the linked documents.

1. Murayama, K., & Elliot, A. J. (2012). The competition–performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin, 138*(6), 1035–1070. — centaur.reading.ac.uk/33980 · record: pubmed.ncbi.nlm.nih.gov/23088570

2. Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin, 136*(3), 422–449. — psycnet.apa.org/record/2010-09071-002

3. Vansteenkiste, M., & Deci, E. L. (2003). Competitively contingent rewards and intrinsic motivation: Can losers remain motivated? *Motivation and Emotion, 27*(4), 273–299. — selfdeterminationtheory.org (PDF) · record: link.springer.com/article/10.1023/A:1026259005264

3a. Senko, C., & Dawson, B. (2017). Performance-approach goal effects depend on how they are defined: Meta-analytic evidence from multiple educational outcomes. *Journal of Educational Psychology, 109*(4), 574–598. — psycnet.apa.org/record/2016-53083-001

4. Card, D., & Giuliano, L. (2016). Can tracking raise the test scores of high-ability minority students? *American Economic Review, 106*(10), 2783–2816. — aeaweb.org (10.1257/aer.20150484) · PDF: davidcard.berkeley.edu (PDF)

5. Murphy, R., & Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies, 87*(6), 2777–2826. — academic.oup.com/restud/article/87/6/2777 · NBER WP: nber.org/papers/w24958

6. Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics, 94*(7–8), 435–452. — sciencedirect.com/.../S0047272710000411 · record: ideas.repec.org

7. Ellison, G., & Swanson, A. (2016). Do schools matter for high math achievement? Evidence from the American Mathematics Competitions. *American Economic Review, 106*(6), 1244–1277. — aeaweb.org (10.1257/aer.20140308) · earlier working paper: nber.org/papers/w18277

8. Ellison, G., & Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives, 24*(2), 109–128. — aeaweb.org (10.1257/jep.24.2.109) · PDF: economics.mit.edu (PDF)

9. Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86*(4), 718–729. — PDF: appstate.edu/~webbrm/jap2001.pdf · record: pubmed.ncbi.nlm.nih.gov/11519655

10. Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science, 24*(5), 648–659. The specific accomplishment statistics quoted in this brief (392 refereed STEM publications; 820 software/patent contributions; 128 creative written works; >1,000 fine-arts achievements; >$25M in grants; ~11.3% holding academic tenure, 7.5% at research-intensive universities; ~44% holding a doctorate in a four-decade follow-up vs. ~2% of the general population) are as reported by Vanderbilt University: news.vanderbilt.edu/2014/07/09/who-rises. SMPY overview & design: my.vanderbilt.edu/smpy

11. Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134*(2), 223–246. — record: pubmed.ncbi.nlm.nih.gov/18298270 · open copy: conservancy.umn.edu (PDF)

12. Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152–161. — sciencedirect.com/.../S0360131514002000

13. OECD (2019). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives* — "Students' self-efficacy and fear of failure." — oecd.org/.../pisa-2018-results-volume-iii (Ch. on self-efficacy & fear of failure)

14. Garcia, S. M., & Tor, A. (2009). The N-effect: More competitors, less competition. *Psychological Science, 20*(7), 871–877. — journals.sagepub.com/.../j.1467-9280.2009.02385.x · record: pubmed.ncbi.nlm.nih.gov/19538434

15. Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest, 12*(1), 3–54. — journals.sagepub.com/doi/10.1177/1529100611418056 · PDF (APA): apa.org/ed/schools/gifted/rethinking-giftedness.pdf

What this brief does *not* establish

- That competition causes learning *on average* — the meta-analytic evidence says it does not [1,11].
- That a specific commercial product improves outcomes — none of these studies tested one; the design section is an evidence-informed argument, not a measured result.
- Magnitudes for any one product or population — effect sizes above are context-specific (e.g., the ~0.5 SD tracking effect [4] or the relative-feedback effect on grades [6] need not transfer). A real product should run its own pilot with retention, learning, equity, and well-being measures (see the MathClash build plan, Phase 0 exit criteria).