EVIDENCE BRIEF · FOR PRINCIPALS, CURRICULUM DIRECTORS AND DISTRICT LEADERS

Coaching that arrives in the classroom while the teacher is speaking to the student.

In-context, in-the-moment professional development — not forgotten in a lecture hall in August.

Teacher Supreme puts research-backed coaching language into a teacher's hand *while the moment is still happening* — then keeps a private, FERPA-safe record of what was said and to whom. This brief states what the research supports, what it does not, and how a school can fund and evaluate it under ESSA.

teachersupreme.com / science

Created by James C. Morris, M.Ed. — a San Diego public school teacher for 35 years, now retired

READ THIS FIRST

There is no study of Teacher Supreme. No randomized trial, no quasi-experiment, no efficacy claim. What follows is the published research the product is *built from*, and an honest account of the distance between that research and this software. Under ESSA that distance has a name and a legal home — **Tier IV, "Demonstrates a Rationale"** — and §5 shows exactly how to document it.

1. The problem you are already paying for

Professional development is not under-funded. TNTP's *The Mirage* studied three large districts and one charter network and put the spend at roughly **$18,000 per teacher per year**, consuming about **19 school days** — a tenth of the year. Across the fifty largest districts, an estimated **$8 billion annually**. The study found no link between those programs and improved classroom performance.[1]

METRIC	DETAIL	SOURCE
$18,000	Spent per teacher, per year, on professional development in the districts studied	TNTP (2015)
19 days	Of the school year given to PD — about 10% of instructional time	TNTP (2015)
$8 billion	Estimated annual PD spend across the 50 largest U.S. districts	TNTP (2015)

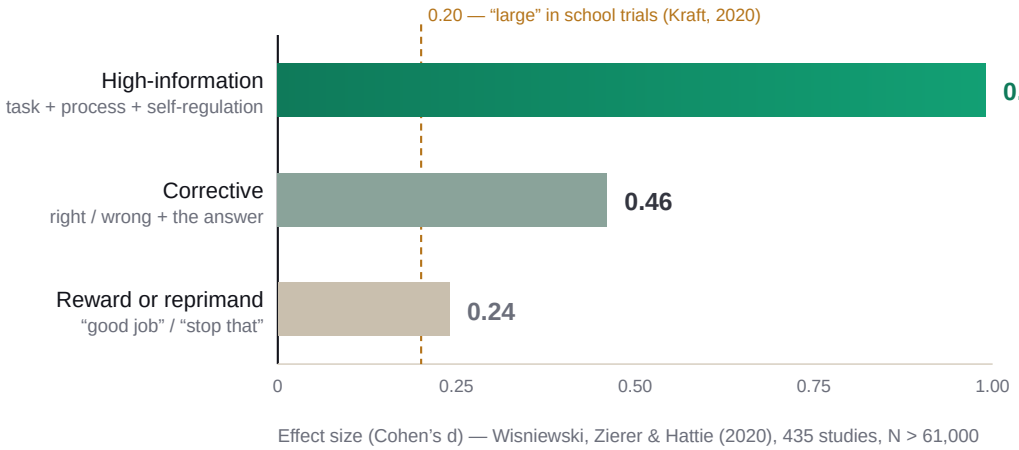
Figures are from a 2015 study of four systems in 2015 dollars. They are the best-known numbers in the field, not a national census, and we present them as such.

The failure mode is not effort. It is **timing**. A workshop in August is asked to change a decision a teacher makes in four seconds in November, without a single reminder in between. The gap between the training room and the moment of practice is where the money goes.

> *A student looks right at you and says "I don't know." You have about four seconds to say something useful. Nobody alive can produce the right sentence on the fly, thirty-six times a period, six periods a day. That is not a skill problem. That is a cognitive load problem.*
>
> — James C. Morris, M.Ed., founder; 35 years teaching in San Diego public schools, now retired

2. What the research says a teacher should say

The single largest finding behind this product is not *that* feedback works — it is that **the kind of feedback decides everything**. Wisniewski, Zierer & Hattie (2020) pooled 435 studies and more than 61,000 participants. Feedback overall averaged $d = 0.48$. But the average hides the story.[2]



The same teacher, the same student, four times the effect. Feedback that names the strategy and hands the student a way to check their own work reaches $d = 0.99$. Feedback that only rewards or reprimands sits at 0.24 — barely above the bar Kraft (2020) sets for a large effect in a randomized school trial. The difference is not effort or care. It is *wording*, produced in about four seconds, thirty times an hour.[2,3]

> **THE MECHANISM, IN ONE SENTENCE**
>
> Teachers already know feedback matters. What no human can do unaided is compose *high-information* feedback — name the move, name the strategy, name the next occasion — thirty times an hour while also teaching. **That composition is the entire job of this software.**

3. Why the exact words matter — and why "good job" is not neutral

Mueller & Dweck (1998) ran six studies with 412 fifth graders. Every child was told the same thing about their score. Then one group heard six extra words — "you must be smart at these" — and the other heard "you must have worked hard at these." That was the whole difference.[4]

AFTER ONE SENTENCE OF PRAISE

“You must be smart”

67%

chose the easier task — the one that would make them look good

After failure: quit sooner, enjoyed it less, scored **below their own first round**

“You must have worked hard”

Harder

chose the task they would learn more from

After failure: stayed with it, blamed effort, finished **higher than they started**

Mueller & Dweck (1998), six studies, 412 fifth graders. Three of the six samples were majority African American and Hispanic.

In one of the six studies, **38%** of the children praised for intelligence misreported their score to children they had never met, against 13% of the effort-praised group and 14% of controls. Person praise hands a student something to protect. Process praise hands a student something to repeat.[4]

This is the rule the product is written to, and it is enforced in the code rather than merely intended. Every coaching line names what the student did, names the strategy, and names the next occasion to use it — addressed to that student by name. Generic praise is not offered.

4. What a teacher is actually holding

Cowan (2001), reviewing decades of experiments, puts working memory nearer **four chunks** than Miller’s famous seven. Cognitive-load theory supplies the consequence: capacity spent on information extraneous to the task is capacity not available for the task.[5]

No study measures a teacher carrying thirty-six students’ allergies, IEP pages, missing work and who has not spoken in a week *while also teaching*. **That extension is ours, not a finding.** But the direction is not in dispute, and it is the reason the product is built the way it is: the teacher does the noticing and the judging; the software carries the remembering.

THE 24 INDICATORS

Teacher observations map to 24 indicators across four domains — **Cognitive Investment**, **Active Participation**, **Cognitive Withdrawal** and **Off-Task Behavior**, six in each. Eleven carry a published Hattie effect size; the other thirteen rest on named observational frameworks and behavioural-science standards and carry none. **Every effect size below is an association, not a promised gain** — it comes from research measuring whether two things travel together, not from trials that assigned some students a better classroom.

The eleven indicators carrying a published Hattie effect size. The remaining thirteen — task avoidance, rapid guessing, learned helplessness, off-task verbal, digital distraction, noncompliance and others — rest on named observational systems (BOSS; Wise & Kong, 2005; Dweck & Reppucci, 1973; Ravizza et al., 2017) and carry *no* published effect size. We list them as such rather than borrowing a number.

WHERE WE DISCLOSE THE DISAGREEMENT

Hattie's effect-size rankings have real methodological criticism in the literature, and his figures are drawn largely from correlational syntheses. They cannot be lined up against Kraft's 0.20 bar, which is set for randomized trials measured on standardized tests. Any vendor presenting Hattie numbers as guaranteed gains is misreading them — and so would we be.

WHAT ONE COACHING MOMENT ACTUALLY CONTAINS

This is not a summary. It is the library entry for *Explains Reasoning*, word for word as a teacher sees it, with the research level of each line labelled. Hattie & Timperley (2007) define three: **task** fixes the work, **process** names the strategy, **self-regulation** hands the student the checking.

EXPLAINS REASONING — ELABORATION & ORGANIZATION · d ≈ 0.7 · HATTIE

TRY FIRST · PROCESS

"You explained every step instead of jumping to the answer. That is showing your reasoning — show your steps again whenever a problem gets hard."

Mueller & Dweck (1998) — process praise: name the move, name the next occasion

THEN · PROCESS

"That is one route to the answer. Now show me a second way to reach the same answer."
The student rebuilds the solution instead of reciting the first one from memory.

Chi et al. (1994) — self-explanation: a second route cannot be pulled from memory

GO DEEPER · PROCESS

Send the student to teach a partner — explain each step and why it works.
The steps they skip or fumble are the ones they only half-knew.

Rosenshine (2012) — explaining aloud forces every step into words

Twenty-four indicators, ninety-nine ranked moves, every one carrying the name of the researcher it came from. **In class nobody reads any of this.** The teacher says what they saw; the line arrives on the student's card while the moment is still happening. The library is browsable by every teacher; the live coaching moment is the paid feature.

WHY THIS MATTERS TO AN INSTRUCTIONAL LEADER

A workshop teaches a teacher *about* Hattie & Timperley's three levels of feedback. This hands them a line at each level, in their own classroom, attached to the student in front of them — and the escalation is built in, so a teacher who uses only the first line is still giving process feedback rather than "good job."

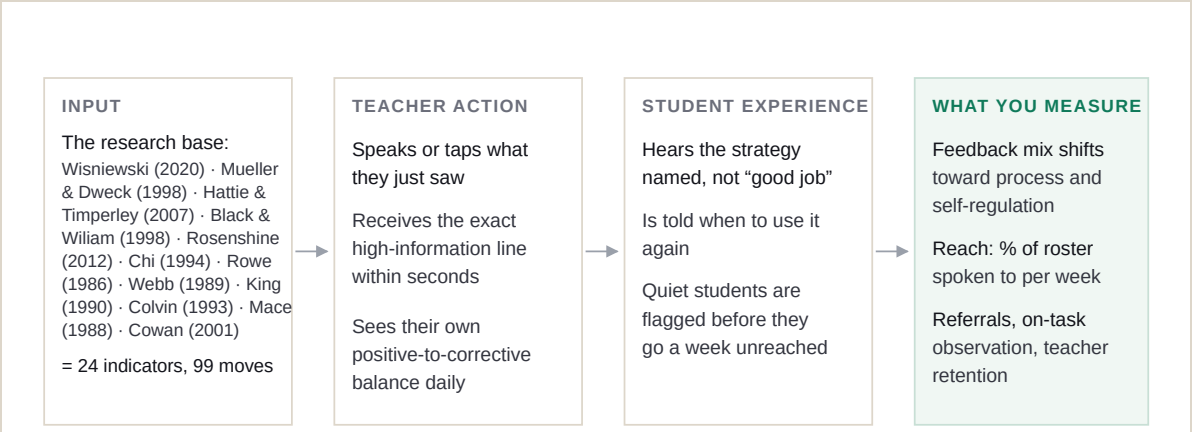
5. How to fund it: ESSA Tier IV, documented

ESSA defines four evidence levels. Teacher Supreme sits at **Level 4** and says so plainly.[6]

LEVEL	WHAT ESSA REQUIRES	TEACHER SUPREME
1 — Strong	At least one well-designed, well-implemented *experimental* study	No. No RCT exists.
2 — Moderate	At least one well-designed *quasi-experimental* study	No.
3 — Promising	At least one well-designed *correlational* study with controls	No.
4 — Rationale	"Demonstrates a rationale based on high-quality research findings… that such activity, strategy or intervention is likely to improve student outcomes" — supported by a **logic model** and a **plan to evaluate**	**Yes — and documented below.**

Level 4 is not a consolation prize; it is the tier written for interventions grounded in strong research that have not yet been trialled as products. It carries a real obligation, and the obligation is the point: explore the research, state the mechanism, and plan the evaluation.[6]

THE LOGIC MODEL



An ESSA Level 4 submission needs exactly this: the research, the mechanism, and the measurement. A school can paste this diagram into its plan.

1. **Baseline, two weeks.** Record each participating teacher's positive-to-corrective ratio and weekly student-reach percentage. Both are produced by the app itself and belong to the teacher.
2. **Compare against a matched group.** Same grade band, same subject, no app. You now have a quasi-experimental design — the standard for ESSA Level 2 if you later choose to submit it.
3. **Track three outcomes at 30, 60 and 90 days.** Office discipline referrals; percentage of roster receiving at least one documented positive per week; teacher-reported time spent on documentation.
4. **Report honestly, including nulls.** We will publish the result whichever way it falls, and name the school only with written permission.

> **FUNDING**
>
> Because it is **professional learning** rather than curriculum, Teacher Supreme is generally considered under **ESSA Title II, Part A** (Supporting Effective Instruction) and, where applicable, Title IV-A. Confirm allowability with your own federal programs office; a Level 4 rationale plus the evaluation plan above is the documentation they will ask for.

6. The privacy answer, before you ask

School people ask this first and should. **Student names are replaced with tokens — STU-001 — on the teacher's own device, before anything is transmitted.** The AI drafts using the tokens; the real names are restored only on the teacher's screen. The transcript visible on the device is already tokenized, so a glance from across the desk shows nothing.

The seating chart and its faces never leave the phone. One tap exports the roster and every logged interaction, because software that makes leaving difficult is telling you something. We describe the mechanism rather than claiming a certification: **we do not say "FERPA certified."** We say what the software is physically handed, and invite you to verify it.

7. What we are not claiming

- No efficacy study of Teacher Supreme exists. Nothing here should be read as one.
- No effect size on this page is a gain any teacher is promised. They are associations from the literature the product draws on.
- Hattie's rankings are contested and largely correlational; his numbers are not comparable to Kraft's randomized-trial benchmark.
- The cognitive-load argument in §4 is an extension we make, clearly labelled, not a published finding about teachers.
- We do not claim FERPA certification, and no such certification exists to hold.

Every one of these disclosures costs us something in a sales conversation. They are here because a research brief that cannot be checked is a brochure.

8. The ask

1. **Pilot with five to ten teachers for 90 days**, using the evaluation above. Every teacher starts on a 28-day free trial; no card is needed to begin.
2. **Review the teachers' own Rounds reports at day 45** — shared voluntarily by the teacher. Rounds are the teacher's private growth tool and are never uploaded to an administrator by the software.
3. **Decide at day 90 on the three outcomes you chose**, not on a vendor's testimonial.

Sources. Every figure in this brief was verified against the primary source on 2026-08-10.

1. TNTP (2015), *The Mirage: Confronting the Hard Truth About Our Quest for Teacher Development* — tntp.org/publication/the-mirage-confronting-the-truth-about-our-quest-for-teacher-development/
2. Wisniewski, B., Zierer, K., & Hattie, J. (2020), "The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research," *Frontiers in Psychology* 10:3087 — doi.org/10.3389/fpsyg.2019.03087
3. Kraft, M. A. (2020), "Interpreting Effect Sizes of Education Interventions," *Educational Researcher* 49(4), 241–253 — doi.org/10.3102/0013189X20912798. Benchmarks for causal studies of standardized achievement: small < 0.05; medium 0.05 to < 0.20; large ≥ 0.20.
4. Mueller, C. M., & Dweck, C. S. (1998), *Journal of Personality and Social Psychology* 75(1), 33–52 — doi.org/10.1037/0022-3514.75.1.33
5. Cowan, N. (2001), "The magical number 4 in short-term memory," *Behavioral and Brain Sciences* 24(1), 87–114 — doi.org/10.1017/S0140525X01003922; Paas & van Merriënboer (2020), *Current Directions in Psychological Science* 29(4), 394–398
6. ESSA §8101(21)(A) evidence definitions; U.S. Department of Education non-regulatory guidance, *Using Evidence to Strengthen Education Investments*. Level 4 requires a rationale grounded in high-quality research, a logic model, and a plan to evaluate.
7. Hattie, J., & Timperley, H. (2007), "The Power of Feedback," *Review of Educational Research* 77(1), 81–112; Black, P., & Wiliam, D. (1998), *Phi Delta Kappan* 80(2)
8. Cook et al. (2017) and Caldarella et al. (2023), *Journal of Positive Behavior Interventions*; Sabey, Charlton & Charlton (2019), *Journal of Emotional and Behavioral Disorders* — which found no empirical support for any specific praise-to-reprimand ratio.

Teacher Supreme · Agentic Edge Ventures · teachersupreme.com · james@agenticedgeventures.com · The full research page, with clickable primary sources, is at teachersupreme.com/science