

The Effect Size Sanity Check

Five questions to run any effect size through before it reaches a slide, a staff meeting or a board packet. Built for grades 6–12 educators, no statistics training required.

An effect size is a translation of a relationship onto a common scale. It is not a measure of whether one thing caused the other — and that difference is where most education research goes wrong on its way to a slide.

#	The question	Why it catches what it catches
1	Correlational or experimental?	If nobody was assigned to anything, the number describes a relationship. Read it as “these travel together,” never “this causes that.” Correlational effect sizes run higher than those from randomised trials.
2	Which direction would common sense run?	Say it out loud. If the reverse story is at least as plausible — success building belief rather than belief building success — that belongs in the sentence where you cite the number.
3	Who were the participants?	A finding from nine-year-olds is not a finding about your sophomores. This question disqualifies more education research than any other on the list.
4	Compared against what?	An effect size is always a comparison. Against a strong alternative, 0.20 is excellent. Against nothing at all, 1.00 is unremarkable.
5	What would have to be true for this to be wrong?	If you cannot answer, you have finished reading a summary of the study rather than the study.

Two benchmarks, and which one applies

	Hattie's hinge point	Kraft's causal benchmarks
The line	0.40 and above is “worth having”	Below 0.05 small · 0.05–0.20 medium · above 0.20 large
Drawn from	An average across a database of meta-analyses, most of them correlational	Nearly 2,000 effect sizes from randomised controlled trials
The catch	Only about 13 percent of randomised-trial effect sizes reach 0.40 at all	Applies to causal studies only — do not use it to judge a correlation

Hattie's own reply, published alongside Kraft's critique, is that 0.40 was offered as an average and a starting point for discussion rather than a universal pass mark. That is a fair account of what he wrote. It is not a description of how the number usually gets used in a faculty meeting.

Three phrases that should make you stop and check

“Research shows...”	Which research? One study, a meta-analysis, or a summary of a summary? The phrase is doing the work a citation should be doing.
“X times more powerful than Y”	Almost always two different kinds of measurement ranked against each other. Ask what each one was measured with before accepting the ratio.
“The research is clear”	Sometimes true. More often it means the speaker has read one side. Ask what the strongest objection is — a good source has one ready.

Run One Yourself

Take the next effect size somebody puts in front of you and work it through. The worked example is the one this sheet came from.

Worked example — collective teacher efficacy, 1.57

1. Correlational or experimental?	Correlational. Belief was surveyed, achievement was measured, and the two were found to travel together. FAILS
2. Which direction?	Genuinely unclear, and the reverse is plausible — a staff whose students have improved for three years has good reason to believe in itself. WOBBLES
3. Who were the participants?	Teachers and students in schools. PASSES
4. Compared against what?	Prior achievement 0.65, socioeconomic status 0.52, home environment 0.52 — all stated. PASSES
5. What would make it wrong?	Rarely asked. If achievement drives belief rather than the reverse, the practical advice changes completely.

Verdict: a strong correlation, reported honestly by its authors, that does not support “believe harder as a staff and scores follow.” That is still a useful finding. It is just a smaller one than the slide implies.

Now yours

The claim, and where you heard it

Write it in the words it was said to you.

#	The question	My answer	Pass / fail

What I would say about this number if I had to present it honestly

One sentence. If it needs three, the number is doing less work than it looked like.

Nothing on this sheet is an argument against using research. It is an argument for citing it at the size it actually is — which is the only way a staff keeps trusting the next number you bring them.