

Every Pedagogy Works in the Right Classroom and No Pedagogy Works in All of Them: A Conditions-Adjusted Meta-Analysis of Educational Interventions and the Teachers Who Were Never Asked

Dr. Margaret Osei-Boateng¹, Prof. David Winterbottom², Dr. Yuki Tanaka-Griffiths³

¹ Department of What Actually Happens in Classrooms, University of Lived Experience, Accra ² Faculty of Interventions That Looked Good on Paper, Canterbury Institute of Optimism ³ Centre for Studies That Would Have Worked If the Students Had Read the Brief, Kyoto-Bristol

DOI: 10.1234/taf.2026.askanytheteacher.0071

Abstract

We examined 11,847 studies of educational interventions published between 1965 and 2025, covering direct instruction, discovery learning, one-to-one tutoring, group work, flipped classrooms, unflipped classrooms, and classrooms that were never told they were part of an experiment in the first place. Our principal finding is that every pedagogical approach works — in the right classroom, with the right teacher, for the right students, at the right time, under conditions that the original researchers either carefully documented or cheerfully ignored. No approach worked universally. The single strongest predictor of student outcomes was not the intervention itself, but what we term the "Thousand Small Things" variable: the accumulated professional judgment, relationship quality, and contextual awareness of the teacher delivering it. We note that this variable does not fit neatly into a regression, which may explain why it has been so comprehensively overlooked.

Keywords: pedagogy, context, teacher judgment, things Ofsted cannot measure, the bleeding obvious

1. Introduction

The history of educational research can be told as a love story between policymakers and silver bullets. In each decade, a new intervention arrives with transformative promise: phonics will save reading, group work will save collaboration, technology will save everything, and class size reduction will save the children — provided someone can find the funding, which nobody can. Each intervention is launched with a landmark study, implemented with varying fidelity, evaluated with contradictory results, and ultimately replaced by the next intervention, which arrives with a fresh landmark study and identical confidence.

Meanwhile, in staffrooms across the world, teachers have been quietly getting on with the work of teaching, which involves making approximately 1,500 micro-decisions per day about when to explain, when to question, when to wait, and when a child needs a glass of water more than they need a learning objective. None of these decisions appear in the methodology section of any randomised controlled trial. This paper investigates whether that omission matters. (It does.)

2. Methodology

We reviewed 11,847 peer-reviewed studies of educational interventions spanning sixty years and covering: direct instruction, inquiry-based learning, project-based learning, one-to-one tutoring, small-group tutoring, peer tutoring, reciprocal teaching, Socratic questioning, flipped classrooms, blended learning, mastery learning, setting by ability, mixed-ability grouping, homework, no homework, more homework, less homework, and one study from 2014 that investigated whether students learn better when the classroom smells of peppermint (inconclusive, though attendance improved).

Each study was coded for: (1) Did the intervention produce positive outcomes? (2) Did it produce positive outcomes every time it was tried? (3) Did the authors acknowledge the role of the teacher, or did they write as though the intervention operated autonomously, like a particularly well-designed thermostat? Coding was performed by nine

research assistants and three former teachers. The former teachers finished first. They also cried less.

3. Results

Table 1 summarises the findings. Of the 11,847 interventions studied, every single one produced measurable positive effects in at least one context. Not a single one produced positive effects in every context. Direct instruction worked brilliantly — when the teacher was clear, the content was well-sequenced, and the students were ready to receive it. Discovery learning worked brilliantly — when the teacher scaffolded it carefully, the resources were right, and the students weren't simply discovering that they were confused. One-to-one tutoring produced the largest average effect sizes of any intervention, which surprised nobody who has ever sat next to a struggling student and watched the moment the penny drops.

Table 1. Summary of findings across 11,847 studies (1965–2025). All results statistically significant. None pedagogically surprising.

Finding	Percentage	p-value
Interventions that worked at least once	100%	< 0.001
Interventions that worked every time	0%	< 0.001
Studies where teacher quality was the dominant variable	100%	< 0.001
Studies that acknowledged this in their conclusions	4.7%	< 0.001
Interventions promoted with "clear implications for practice"	91.3%	< 0.001
Teachers who were consulted before implementation	6.1%	0.003
Studies involving Friday Period 5	0.4%	N/A (survival bias)

The most significant finding concerned what we have called the "Thousand Small Things" variable — a composite measure of teacher expertise, relational awareness, adaptive decision-making, and the kind of professional judgment that develops over years and cannot be captured in a CPD workshop titled "Transforming Your Practice in 90 Minutes." When this variable was included in the model, it accounted for between 61% and 94% of the variance in student outcomes, reducing the independent contribution of the named intervention to what one reviewer called "a rounding error with a marketing budget."

We also found a strong inverse correlation ($r = -0.81$, $p < 0.001$) between the confidence with which an intervention was promoted by people outside the classroom and its reliability when implemented by people inside one. The more certain the policy brief, the less certain the classroom outcome. We have tentatively named this the "Keynote Speaker Effect."

4. Discussion

These findings pose an uncomfortable question for anyone who has ever stood at a lectern and told a roomful of teachers what "the evidence says." The evidence, it turns out, says that everything works sometimes and nothing works always, and the difference is almost entirely determined by the skill, judgment, and relationships of the person delivering the intervention — a person who was typically not consulted when the intervention was designed, selected, or mandated.

We note that the phrase "context-dependent" appeared in only 4.7% of study conclusions. The phrase "these findings have clear implications for policy and practice" appeared in 91.3%. We leave readers to judge which phrase more accurately reflects the state of human knowledge about why some children learn long division on Tuesday and the same children cannot remember it on Thursday.

It is also worth observing that the most effective teachers in our sample did not align with any single pedagogical philosophy. They used direct instruction when students needed clarity. They used inquiry when students were ready to explore. They told jokes when the energy dipped. They adjusted when it wasn't working, which is to say, they did the thing that professional expertise consists of but that no intervention manual has yet found a way to mandate.

5. Conclusion

Every pedagogy works. No pedagogy works everywhere. The difference is the teacher, the students, the Tuesday, and the thousand things that never make it into a dataset. We recommend that future educational research be required to include a section titled "What the Teacher in Room 4B Would Say About This," though we acknowledge this section would likely be more useful and considerably shorter than the rest of the paper. We further recommend that policymakers considering the next transformative intervention first ask the question that all evidence, across six decades, converges on: "Who is going to teach it, and do they think it will work?" The answer, like everything else, will depend on context.

Acknowledgements

The authors wish to thank the three former teachers on our research team, whose contributions were invaluable and whose salary expectations were, by academic standards, disarmingly modest. We are grateful to Reviewer 2, who wrote that our paper was "a 4,000-word way of saying what every teacher already knows," which we believe was intended as a criticism but which we accept as a precise summary of our contribution. We also thank the students of Year 9 Set 3, period 5, on a Friday afternoon, whose existence disproves every educational theory simultaneously.

References

- Bloom, B. S. (1984). "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." *Educational Researcher*, 13(6), pp. 4–16. [Note: still searching.]
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge. [Reviewed by 10,000 policymakers, understood by considerably fewer.]
- Osei-Boateng, M. (2024). "The Staffroom Knows: Practitioner Wisdom as an Uncontrolled Variable." *Journal of Things We Should Have Measured*, 12(1), pp. 33–51.
- Tanaka-Griffiths, Y. (2022). "Friday Period 5: A Natural Experiment in the Limits of Pedagogical Theory." *Kyoto-Bristol Papers on Educational Realism*, 8(2), pp. 77–89.
- Winterbottom, D. (2023). "The Keynote Speaker Effect: Confidence as an Inverse Predictor of Classroom Reliability." *Canterbury Review of Evidence-Adjacent Claims*, 19(4), pp. 201–218.
- Various (1965–2025). "This Approach Has Clear Implications for Practice." *Every Education Journal Ever Published*, passim.

Disclaimer: The Appropriate Forum is a satirical publication. No pedagogies were harmed in the production of this study, though several were asked to leave the room and reflect on their choices.

© 2026 The Appropriate Forum. All rights reserved, though the authors acknowledge that the enforceability of this claim depends, like all things, on context — and on who is teaching it.