

PROFESSIONELLE LERNGEMEINSCHAFT SCHULPRAKTIKUM III

2025, Nikolaus Albrecht

Skript: <http://www.sigmadelta.at/>

INHALT:

- 1_ Organisation, Ablauf und was Sie erwartet
- 2_ What makes great teaching | Robert Coe
- 3_ Effect size
- 4_ Misrepresenting B. Bloom & the learning pyramid

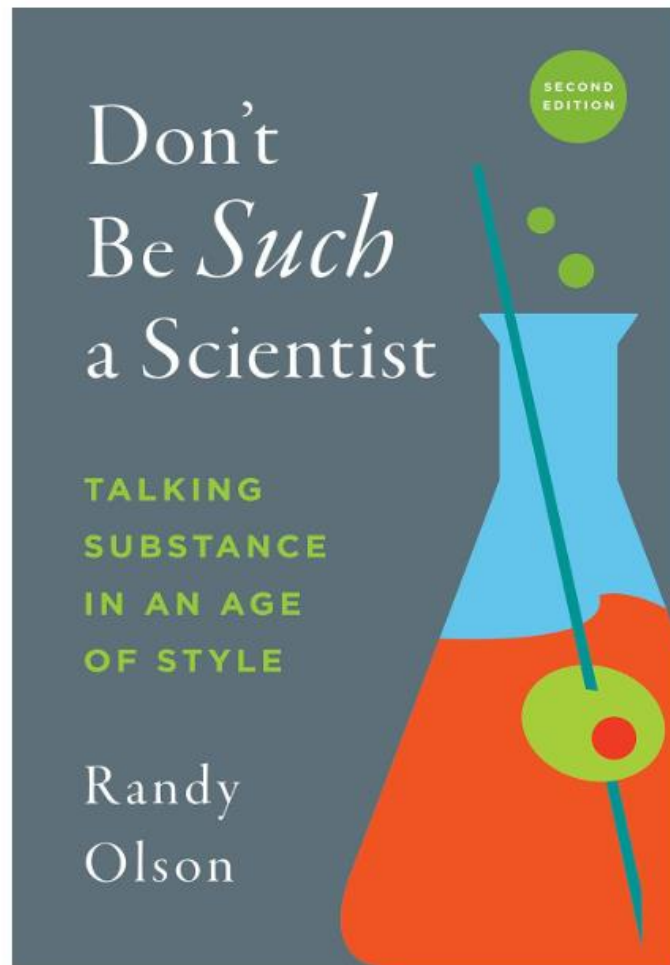
1_Organisation, Ablauf und was Sie erwartet

- die ersten 1,5 h synchron online
- die letzten 0,5 h asynchron (Lektüre, Vertiefung,...)
- bei Bedarf kann natürlich davon abweichend gearbeitet werden

#_Rückblende 30/9/2025

Die Entdeckung der „Grautöne“ für die Lehre und ein Versuch der Vermeidung von einfacher Schwarz – Weiß-Malerei in der Didaktik

Lesetipp:



2_What makes great teaching | Robert Coe

<https://eric.ed.gov/?id=ED559263>

<https://www.suttontrust.com/our-research/great-teaching/>

What makes great teaching?

Robert Coe,

200 pieces of research

Many popular teaching practices are ineffective

Key Findings

The two factors with the strongest evidence of improving pupil attainment are:

- teachers' content knowledge, including their ability to understand how students think about a subject and identify common misconceptions
- quality of instruction, which includes using strategies like effective questioning and the use of assessment

Specific practices which have good evidence of improving attainment include:

- challenging students to identify the reason why an activity is taking place in the lesson
- asking a large number of questions and checking the responses of all students
- spacing-out study or practice on a given topic, with gaps in between for forgetting
- making students take tests or generate answers, even before they have been taught the material

Common practices which are not supported by evidence include:

- using praise lavishly
- allowing learners to discover key ideas by themselves
- grouping students by ability
- presenting information to students based on their “preferred learning style”

Quiz on Research-Based Pedagogy

<https://www.cambridgeinternational.org/Images/557877-rob-coe-what-makes-great-teaching-.pdf>

Class size

1. Reductions in class size (eg 30→20) generate
 - a) Substantial increases in students' learning
 - b) Small increases in students' learning
 - c) No change in students' learning

<https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit>

→ Class size

John Hattie

https://inspirasifoundation.org/wp-content/uploads/2020/05/John-Hattie-Visible-Learning_-A-synthesis-of-over-800-meta-analyses-relating-to-achievement-2008.pdf

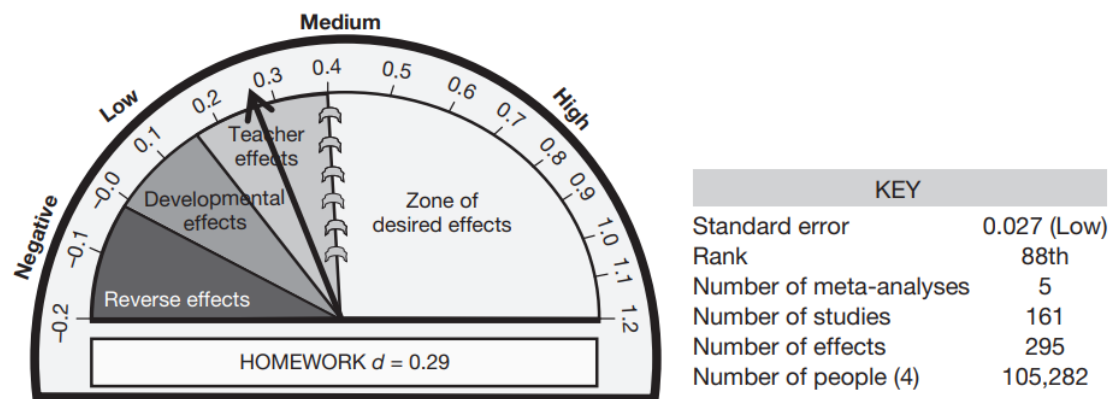
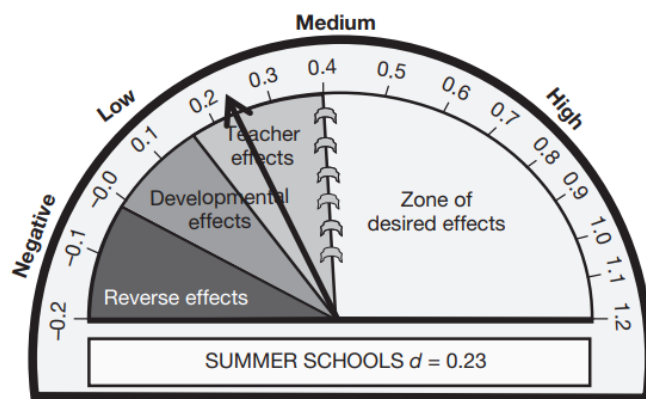
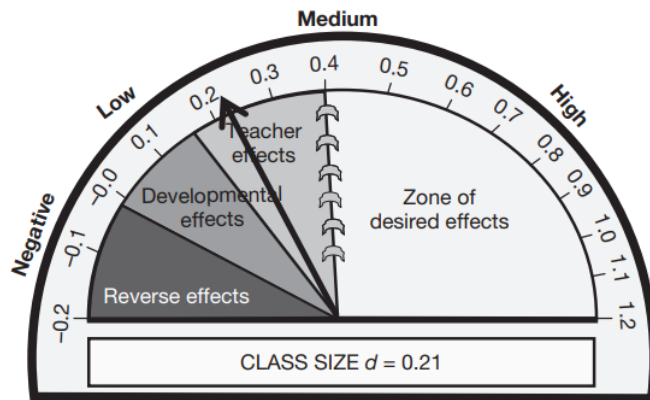


Figure 2.4 A typical barometer of influence

4	The contributions from the student	39
5	The contributions from the home	61
6	The contributions from the school	72
7	The contributions from the teacher	108
8	The contributions from the curricula	129
9	The contributions from teaching approaches—part I	161
10	The contributions from teaching approaches—part II	200



KEY	
Standard error	na
Rank	98th
Number of meta-analyses	3
Number of studies	105
Number of effects	600
Number of people (2)	28,700



KEY	
Standard error	na
Rank	106th
Number of meta-analyses	3
Number of studies	96
Number of effects	785
Number of people (x)	550,339

<https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/reducing-class-size>

Behind the average



Impact on mathematics is higher (+2 months) than reading (+1 month).



Most studies examine reductions of 8–10 pupils. The impact of studies that examine reducing class sizes by 5 pupils is smaller, on average.

@ Class size:

Closing the disadvantage gap

International research evidence suggests that reducing class size can have positive impacts on pupil outcomes when implemented with socioeconomically disadvantaged pupil populations. Some studies also have also found that smaller class sizes in primary schools can have a greater positive impact on disadvantaged pupils than their peers.

How secure is the evidence?

The security of the evidence around reducing class size is rated as very limited. 63 studies were identified. The topic lost padlocks because:

- A large percentage of the studies are not randomised controlled trials. While other study designs still give important information about effectiveness of approaches, there is a risk that results are influenced by unknown factors that are not part of the intervention.
- A large percentage of the studies were not independently evaluated. Evaluations conducted by organisations connected with the approach – for example, commercial providers, typically have larger impacts, which may influence the overall impact of the strand.
- There is a large amount of unexplained variation between the results included in the topic. All reviews contain some variation in results, which is why it is important to look behind the average. Unexplained variation (or heterogeneity) reduces our certainty in the results in ways that we have been unable to test by looking at how context, methodology or approach is

Evidence strength ?



Number of studies

63

Review last updated

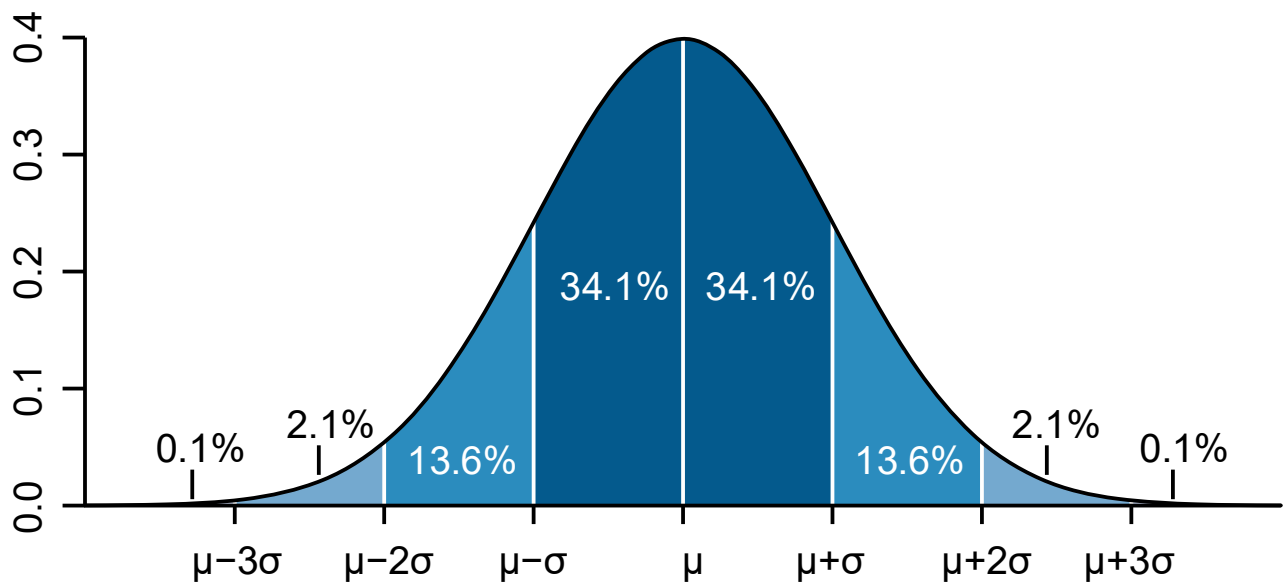
October 2025

The evidence suggests that significant effects of reducing class size are not seen until the number of pupils has decreased substantial (to fewer than 20 or even 15 pupils). Crucially, a reduction in class size is only likely to be effective if it permits teachers to change their teaching approach to the extent that this changes the learning behaviours of pupils. High quality implementation of reducing class size might consider:

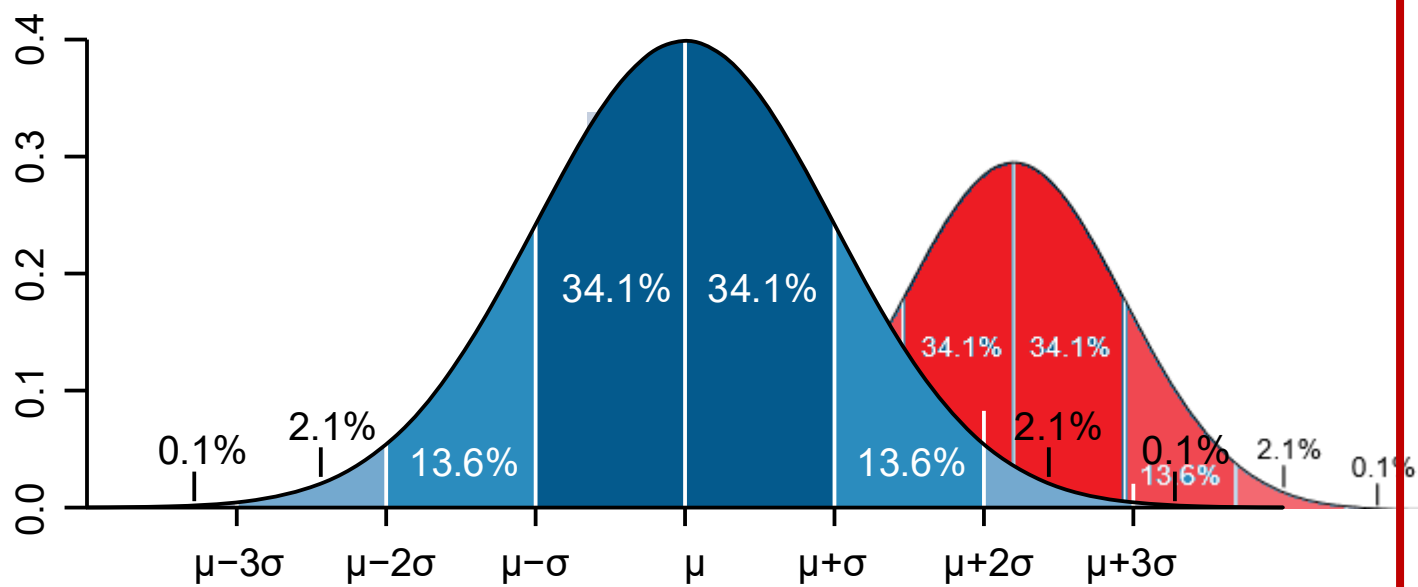
- Additional opportunities to provide feedback on pupils
- Time for high quality interaction between pupils and teachers e.g. modelling approaches closely with pupils.

Bloom, B: “The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring” EDUCATIONAL RESEARCHER, (1984)

<https://web.mit.edu/5.95/www/readings/bloom-two-sigma.pdf>



By Ainali - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=3141713>



By Ainali - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=3141713>

Takeaways:

Most students can achieve high results given the right instructional approach.

Effect size

In educational (and broader social science) research, when people talk about *effect size* d , they almost always mean **Cohen's d** , which is one of the most widely used standardized effect size measures.

Here's the definition:

$$d = \frac{M_1 - M_2}{SD_{\text{pooled}}}$$

- M_1, M_2 : the means of the two groups being compared (e.g., experimental vs. control group).
- SD_{pooled} : the pooled standard deviation, i.e. a weighted average of the two groups' standard deviations.

The pooled SD is usually computed as:

$$SD_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

where n_1, n_2 are the group sample sizes and SD_1, SD_2 are their standard deviations.

Bloom:

“... many teachers are only getting feedback on what students are learning from a small sample of high achievers in the class, usually the ones who raise their hands.”

➔ Cold Call

18.11

THE LEARNING DISPATCH

EXPLORING THE SCIENCE OF LEARNING

The Algorithmic Turn: The Emerging Evidence On AI Tutoring That's Hard to Ignore

Are We Approaching A Turing Test for Teaching?

CARL HENDRICK

NOV 08, 2025

<https://carlhendrick.substack.com/p/the-algorithmic-turn-the-emerging>

Emerging research in AI-tutoring – Hendrick reviews recent trials showing that **algorithm-driven tutoring systems (powered by large-language models / AI) are beginning to show performance gains over more traditional instructional approaches in certain contexts.**

He raises the question: if tutoring is algorithmic, then how uniquely human is teaching?

The optimist ...

The optimistic vision is compelling: every child receives expert, tireless, infinitely patient instruction **calibrated precisely to their needs**. The achievement gap narrows because the students who most need help finally get it, not in sporadic bursts but continuously, systematically.

AI doesn't get tired, it doesn't lose focus, it doesn't have to manage thirty students at once. In theory, it delivers feedback at the exact moment it's needed, never too late, never too soon. It **adjusts the pacing** not to the median of a class but to the learner's individual rate of forgetting. **It never forgets what the student has mastered or misunderstood.**

Taken together, the findings from AI tutoring point to a pattern that echoes what learning-science has been telling us for decades:

The systems are simply better at applying the known laws of learning, ones we have known for 100 years; **explicit instruction, timely feedback, discriminating between varied examples, adaptive pacing, retrieval practice spaced out, and integrating new knowledge with old.**

The optimist can point to a set of studies ...

<https://www.nature.com/articles/s41598-025-97652-6>

scientific reports

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 03 June 2025

AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting

[Greg Kestin](#) , [Kelly Miller](#), [Anna Kiales](#), [Timothy Milbourne](#) & [Gregorio Ponti](#)

[Scientific Reports](#) **15**, Article number: 17458 (2025) | [Cite this article](#)

41k Accesses | **18** Citations | **87** Altmetric | [Metrics](#)

In a randomized trial in a physics course (~200 students), the AI tutor outperformed “in-class active learning” on median learning gains; many students learned faster.

This was not a comparison against passive instruction or weak teaching (as many studies are), but against well-implemented **active learning delivered by highly rated instructors** in a course specifically designed around pedagogical best practices. **The AI tutor produced median learning gains more than double those of the classroom group.**

The effectiveness depended on **careful engineering of the AI - system**. Students cannot simply use ChatGPT or any other off-the-shelf AI tool and expect comparable results.

The system was built by instructors who understood both the content and the **pedagogical principles that promote learning**. This required significant time and expertise.

Perspective ...

What happens when the next generation AI-models can reason through physics problems independently? When they can diagnose misconceptions in real time?

And the Harvard study is not an isolated finding ...

[ASSISTments](#), a mathematics tutoring platform evaluated across two large-scale randomised controlled trials involving thousands of students, achieved effect sizes of 0.18 to 0.29 standard deviations on standardised tests, with the largest gains for struggling students, earning it the highest ESSA Tier 1 evidence rating at a cost of less than £100 per student.

[Carnegie Learning's MATHia](#), tested with over 18,000 students across 147 schools, produced effect sizes ranging from 0.21 to 0.38 standard deviations.

A large randomised controlled trial known as [Tutor CoPilot](#) found that school pupils whose tutors used an AI assistant achieved significantly higher mastery rates than those in the control group, with the biggest gains among the least experienced human tutors.

But pessimists can point to a disturbing truth:

The Illusion of Understanding: When AI Harms Learning

Yet there is a troubling paradox at the heart of AI tutoring. The very same technology that can produce effect sizes above 0.7 standard deviations **can also make students demonstrably worse at learning.**

And Carl Hendrick would argue that **the harmful version is the one most students are currently using today.**

It is in the design ... (of standard solutions)

AI - systems are engineered for user-friendly problem-solving, not for the cognitively effortful process through which understanding is built.

Recent research is beginning to quantify what teachers have long suspected: when AI does the thinking, students stop doing it themselves.

In a 2025 mixed-methods study published in [Societies](#), Michael Gerlich **found that frequent AI tool use was strongly negatively correlated with critical thinking ability**, largely because of a mechanism **known as cognitive offloading**.

Open Access Article

AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking

by Michael Gerlich 

Center for Strategic Corporate Foresight and Sustainability, SBS Swiss Business School, 8302 Kloten-Zurich, Switzerland

“... **creates the illusion of learning**; the sense that one is mastering material **when, in fact, the machine is doing the mastery.**”

And again, this is more than ONE study ...

This is not a peripheral concern. [A rigorous study from the University of Pennsylvania](#) involving high school mathematics students found that **unrestricted access to generative AI without guardrails significantly harmed learning outcomes**. Students with AI access performed worse on subsequent assessments than those who worked through problems unaided. The mechanism is straightforward: when the AI provides solutions on demand, students bypass the very cognitive processes that build understanding. They mistake fluent AI-generated explanations for their own comprehension, a metacognitive error with serious consequences.

The difference lies in the design ...

The distinction between AI systems that enhance learning and those that destroy it is not about the underlying technology.

GPT-4 powered both the highly effective Harvard tutor and the ineffective tools students use to avoid thinking. The difference lies entirely in design.

The Harvard system was engineered to resist the natural tendency of AI - systems to be maximally helpful. It was constrained to scaffold rather than solve, to prompt retrieval rather than provide answers, to increase rather than eliminate cognitive load at the right moments.

ChatGPT, by contrast, is optimised for frictionless task completion. It will happily write your essay, solve your equation, explain the concept you should be puzzling through yourself. It is designed to be helpful, not to promote learning, and **those are fundamentally different objectives.**

Teachers cannot simply “*use AI*”; they must understand the difference between AI as helpful for learning and AI as harmful for learning.

One is a scaffold that can eventually be removed; the other is a crutch that makes walking without it progressively harder.

Teacher expertise is astonishingly complex, tacit, and context-bound. It is learned slowly, through years of accumulated pattern recognition; seeing what a hundred different misunderstandings of the same idea look like, sensing when a student is confused but silent, knowing when to intervene and when to let them struggle. **These are not algorithmic judgements** but deeply embodied ones, the result of thousands of micro-interactions in real classrooms. That kind of expertise doesn't transfer easily; it can't simply be written down in a manual or captured in a training video.

But what AI systems could fundamentally achieve...

What works consistently across the evidence is the combination of immediate feedback, spaced practice, adaptive personalisation, and mastery-based progression.

If we take learning to be a durable change in long-term memory and if we take instruction as the key lever of that and if AI can teach better than humans, not as some distant possibility but as an emerging reality, then we must reckon with what that reveals about teaching itself.

So far, we have only gotten this far...

Many EdTech interventions so far have been solutions in search of problems, designed by technologists with limited understanding of how learning actually occurs. They have prioritised engagement over mastery, confusing students' *enjoyment* of a platform with their acquisition of knowledge. They have ignored decades of cognitive science research in favour of intuitive but ineffective approaches. They have failed to account for implementation challenges, teacher training requirements, and the messy realities of classroom practice.

Perhaps the answer is that teaching and learning are not the same thing, and we've spent too long pretending they are. Learning, the actual cognitive processes by which understanding is built, may indeed follow lawful patterns that can be modelled, optimised, and delivered algorithmically. The science of learning suggests this is largely true: **spacing effects, retrieval practice, cognitive load principles, worked examples; these are mechanisms, and mechanisms can be mechanised. But teaching, in its fullest sense, is about more than optimising cognitive mechanisms.** It is about what we value, who we hope our students become, what kind of intellectual culture we create.

What features of tutoring do you believe are easily algorithmic? What features aren't?

Which tasks could an AI handle? Which tasks must remain human?

What role does the teacher play then?

When we talk about retrieval practice in schools, we're not really talking about retrieval in isolation; we're talking about retrieving knowledge as it occurs in the wild; entangled with motivation, prior knowledge, attention, classroom climate, curriculum sequencing, and the unpredictable dynamics of thirty students learning together.