# Appendix 1: Review methodology

This is a review of existing evidence about what teacher competencies (i.e., teacher behaviours, skills, knowledge, beliefs or other abilities) are 'best bets' to be worth learning in order to improve the impact of teaching. As well as these teacher competencies, we are also interested in evidence about environmental proxies that may provide a valid and timely indicator of the quality of student learning taking place in a classroom. For example, if research showed that relationships of trust and respect between students and teachers were predictive of more learning, and that feeding back an indicator of the quality of those relationships could help to improve them, then we might want to include this in our model, even though it is not directly capturing a specific teacher behaviour.

## Research questions

1. What teacher competencies (i.e., teacher behaviours, skills, knowledge, attitudes, beliefs or other abilities) are 'best bets' for a teacher to try to learn in order to improve the impact of their teaching?
    a. How have these competencies been captured in existing studies? (Are there adequately reliable and valid measures?)
    b. Are they predictive of student learning? What kinds of relationships have been sought/found (e.g., linear, non-linear or threshold effects)?
    c. What do we know about the dependencies among different competencies? Do they interact or depend on each other?
    d. Is there evidence these competencies can be learnt?
    e. Is there evidence that deliberate attempts to improve the competency lead to greater student learning?

2. What kinds of specific, real-time measures of the quality of a classroom environment may be useful, immediate proxies for the student learning that is occurring?
    a. What evidence supports their use as valid indicators of classroom quality?
    b. Is there evidence that feedback to teachers based on these indicators can help them to improve?

# Overview of approach

**Systematic review:**
In formal terms, a systematic review uses a very explicit and precise procedures to identify, select, appraise, and synthesise existing research – sometimes encompasing hundreds or thousands of studies. This section explains how we took a reasoned, practical approach; while valid, it would not fall into the technical category of a systematic review.

The literature that could potentially be relevant to both these questions is so big and diverse that to attempt a comprehensive, **systematic review** would be a colossal task. We certainly did not have the time and resource for such a project, and it is debatable whether this would be time well spent.

Nevertheless, any review of this kind must address two kinds of threats:

- Comprehensiveness: How do we know we have included everything that is relevant? Have we missed or excluded things that should have been considered or included?
- Bias: Might we have emphasised or favoured perspectives or studies that offer a narrow or particular view, perhaps at the expense of other viewpoints?

Our approach was to conduct a rapid 'umbrella' review (i.e., a review of existing reviews), though in many cases we also reviewed the original studies directly, and our searches generated valuable individual studies as well as reviews. From these studies and reviews we extracted a list of the different teacher competencies that have been cited as related to student learning and the environmental proxies that have been claimed as indicators of classroom quality. For each of these elements we evaluated the quality and relevance of evidence supporting its inclusion in a model of 'What is worth learning for teachers?'.

## Identifying relevant studies

We used two main approaches to identifying studies: using known reviews and additional systematic search.

Existing known reviews provide a good place to start in a rapid evidence synthesis. Once we had a list of key reviews, we were able to use backwards (studies they cite) and forwards (later studies that cite them) citation search and related articles search (i.e., studies whose citations overlap). Our starting list included both research reviews and existing frameworks.

1. Reviews:
    a. What makes great teaching? Review of the underpinning research (Coe et al., 2014)
    b. Principles of Instruction (Rosenshine, 2010)
    c. Improving Quality in Education: Dynamic Approaches (Creemers & Kyriakides, 2011)
    d. Effective Teaching: A review of research and evidence (Ko et al., 2013)
    e. State of the art – teacher effectiveness and professional learning (Muijs et al., 2014)
    f. Teacher quality and student achievement (Darling-Hammond, 2000)
    g. Improving students' learning with effective learning techniques (Dunlosky et al., 2013)
    h. Visible Learning for Teachers (Hattie, 2012)

2. Frameworks:
    a. Early Career Framework for England (DfE, 2019)
    b. Enhancing Professional Practice: A Framework for Teaching (Danielson, 2007)
    c. CLASS (Pianta et al., 2012)[3]
    d. ISTOF (Muijs et al., 2018)
    e. ICALT (van de Grift et al., 2017)

In addition, we conducted keyword/topic searches of Web of Science, ERIC and Google Scholar. For Web of Science and ERIC the following string generated 18 and 53 hits, respectively:

("teaching effectiveness" OR "teaching quality" OR "teacher impact" OR "teacher effectiveness" OR "teacher quality" OR "teacher skill" OR "teacher characteristics" OR "pedagogical practice") AND ("learning" OR "attainment" OR "student outcomes") AND ("impact" OR "effect" OR "effects") AND ("systematic review" OR "meta analysis" OR "meta-analysis")

3 See also: **https://curry.virginia.edu/classroom-assessment-scoring-system**

Search strings in Google Scholar are limited to 256 characters and generate thousands of hits, so we used the following string and screened the top 100:

("teaching effectiveness" OR "teaching quality" OR "teacher impact" OR "teacher quality" OR "teacher characteristics") AND ("learning" OR "attainment" OR "student outcomes") AND ("impact" OR "effect") AND ("systematic review" OR "meta analysis")

Results were screened on title and abstract and then obtained and reviewed if they seemed relevant to the research questions above (and had not already been captured from the reviews/frameworks). This was more of an ad-hoc than a systematic process, but it allowed us to check that there were no significant omissions from our evidence base derived from known reviews.

## Extracting information

For each claim in each review or study identified from the search process, we attempted to record:

- Type of design/evidence: theoretical, correlational, interventional, experimental
- Types of student outcome captured (and the quality of measures used)
- Types of teacher competency captured
- Types of environmental indicator captured
- Strength of the relationship found (either conditional or unconditional – specify which and conditioned on what)
- Context of the study: location, date, student age range
- Quality of the study and strength/relevance of the claims

In practice, many of these details were not readily available and the data extraction process was less systematic and thorough than we might have achieved with more time and resource. Nevertheless, we believe we were able to achieve sufficient saturation of findings and a good compromise between comprehensive and manageable.

# Appendix 2: Overview of studies reviewed

## Rosenshine (2010): Principles of Instruction

Rosenshine's (2010) "Principles of Instruction" seems largely to have a similar audience to that of the Great Teaching Toolkit. It focuses on "aspects of language learning and instruction that are universal" and proposes adapting the suggestions to local conditions. The ten principles are derived from three sources:

- Research in cognitive science—how the human brain acquires and uses information, as well as the limits of working memory
- Observation of master teachers—those whose classrooms make the highest gains on achievement tests
- Findings of studies that taught learning strategies

A key consideration for Rosenshine's work is its research base. In the 2010 publication, two "suggested readings" are proffered to further illustrate these claims. However, these pieces of research are not themselves reviews, but more often small-scale, limited interventions or correlational studies. This is not to say that the principles do not come from a large body of literature that supports these practices. If these do exist, however, they are neither directly cited nor signposted. Furthermore, the observational nature of some of the argumentation (e.g., "I once observed a class") potentially belies a systematic, evidence-based argument.

Ultimately, given these reservations, Rosenshine's list reads more as ten specific practices that can be observed in good teachers, rather than broad practices with strong evidence bases. The list comes across as theoretical; it appears to be Rosenshine's (perhaps well-informed) musings. Without greater detail about the outcome measures captured, it remains difficult to further validate his argument. With a focus almost purely on cognitive science, the list does not address any practices pertaining to classroom management, environment, teacher knowledge, etc.

Rosenshine, while presenting principles that on face value seem plausible, leaves a significant gap for offering a more evidence-based argument. Some of his earlier work may offer a more rigorous or systematic approach to the literature (and indeed his somewhat arbitrary selection of "further reading" hints at a deep familiarity with the corpus); however, he has provided no information how earlier conceptualisations and reviews have progressed into these ten principles.

1. Begin a lesson with a short review of previous learning
2. Present new material in small steps with student practice after each step
3. Ask a large number of questions and check the responses of all students
4. Provide models
5. Guide student practice
6. Check for understanding
7. Obtain a high success rate
8. Provide scaffolds for difficult tasks
9. Require and monitor independent practice
10. Engage students in weekly and monthly review

## Muijs et al. (2014): 'State of the Art' review

Muijs et al. (2014) offer a very different sort of review than Rosenshine. Grounded firmly in the "best evidence", they enumerate classroom behaviours that are positively related to student achievement. While the reference list is extensive and filled with recognisable names and studies, the authors do not explain any selection criteria or search methodology to collect these resources; an initial reading suggested a "greatest hits" sort of approach.

The authors highlight six "behaviours" that they argue have the strongest research base. Some of these are focused on cognitive processes (e.g., "Opportunity to learn and time on task" and "Instruction and interaction"), while are some are focused on other aspects of teaching (e.g., "Classroom climate" and "Teacher expectations"). They also highlighted notable meta-analyses that seek to quantify effective teaching strategies.

Because most of the studies discussed in the section on effective behaviours focus on "basic skills in English and mathematics", they also explore significant research into self-regulated learning and non-cognitive outcomes (e.g., wellbeing, self-concept, motivation, etc.).

Additionally, the authors discuss the dynamic model of educational effectiveness. A key feature of the model is that numerous levels have an effect on student achievement. Within the teacher level, they highlight the model's eight factors and associated elements—these are observable instruction behaviours. These elements, as they are presented by Muijs et al., are generally broad approaches (e.g., "Dealing with student responses" and "Promoting the idea of modelling"), with a few more specific behaviours (e.g., "Outlining the content to be covered and signalling transitions between lesson parts" and "Analysing data in order to identify student needs and report the results to students and parents").

The authors also include a section on what these mean for teacher professional development—both its implementation and content. This section seems less relevant to the current work of the Great Teaching Toolkit. The article concludes without a clear direction but with "an invitation to dialogue".

Overall, their work appears to have a strong basis in educational research. They admit that much of the evidence is drawn from research in specific fields and with basic skills, with student achievement as a

typical outcome variable. However, they attempt to address this gap with the second section.

Ultimately, this review has two shortcomings. First, it does not present a single, clear framework that unifies the issues raised. Given the broad approach it had to the behaviours and elements it discussed, it couldn't even be seen as an effective "shopping list" of things to do. Even if they were behaviours or elements that were succinct, the lack of a central organisation leaves the reader unsure of which items to extract. Is it content from the six-widely researched behaviours? The eight factors of the dynamic model? The twenty elements that are embedded within the eight factors?

Second, the article is not accessible to teachers. In a literal sense, it is behind a paywall on the Taylor and Francis website. Furthermore, the register and voice of the writing make it clear that the intended audience is not meant to be teachers. The focus ultimately was for researchers—and the call to action was not for teachers to implement this, but for the research community to engage with the topics at hand.

## Darling-Hammond (2000)

Darling-Hammond's work is often referenced in literature on teacher effectiveness. Her research builds on previous research on individual teachers' attributes, and shifts the focus to a state-level, aggregate focus. Previously, there was little evidence of a relationship between a teacher's academic ability and students' outcomes (as measured by test scores). Somewhat stronger evidence existed of some correlation between a teacher's subject knowledge and student outcomes, but this was only true up to a point—after a certain point, being an expert in an academic field does not translate to increased student learning. However, a stronger-still relationship had been found between a knowledge of teaching and learning and student performance.

Given this body of research, Darling-Hammond conducted a large-scale survey of school characteristics in states across the United States. She also collected state-level data on elementary math and reading achievement. Because the research focuses on aggregate data (i.e., schools and states), the methodology does not focus on classroom practices or techniques. Among the notable findings is a negative correlation between students living in poverty, English as Additional Language learners, and minority ethnicity students and outcomes. Additionally, teacher quality, as indicated by holding a teaching certification and subject degree, has a positive correlation with student outcomes. Two decades later, these findings may not surprise readers, but they were influential in major educational policies in the US.

While it offers us little by way of classroom practices of effective teaching, it further evidences the development of teachers. Teachers can improve their practice; in doing so, it is crucial for them to achieve certain thresholds to progress to higher levels of effectiveness.

## Baumert et al. (2010)

Baumert et al. (2010) examine the concepts of content knowledge and pedagogical content knowledge (PCK) in secondary mathematics teachers in Germany. While previous research had viewed subject knowledge as a unitary concept (e.g., Hill et al., 2004), Baumert et al. explore, both conceptually and empirically, models of the content knowledge and PCK as separate, related concepts. Their assumption is that content knowledge was a sort of threshold prerequisite for PCK, but could not act as a substitute.

The researchers examined a representative sample of Grade 10 mathematics teachers in both the academic and non-academic track in Germany. As an extension to the PISA study, teachers of mathematics answered questionnaires and tests of their teaching knowledge; the questionnaires covered

their background (including training), motivations and beliefs about teaching, and professional beliefs. The tests of teaching knowledge assessed both content knowledge in mathematics and PCK – the latter through open-ended tests of hypothetical situations. Furthermore, teachers submitted homework, tests and classroom tasks to be assessed. Similarly, their students' achievement was measured through tests.

Teachers trained for the academic track show significantly higher scores for their content knowledge and pedagogical content knowledge. This is stronger for content knowledge (greater than one standard deviation) than for PCK; the authors hypothesise this could be because of higher requirements placed on certification, or the higher demands of teachers, on the academic track. These differences remain throughout the teachers' careers. The research finds that the greatest prediction of teachers' content knowledge and PCK scores is the type of teacher training programme attended.

39% of the variance in classes' mathematics achievement (without controlling for the academic or non-academic track) is attributable to pedagogical content knowledge of the teacher. While there are socio-political considerations from the findings, they offer a clear conclusion based on their empirical evidence: teachers' pedagogical content knowledge explains the greatest component of increased student achievement.

- Components of pedagogical content knowledge:
  - Tasks – teachers' ability to identify multiple solution paths
  - Students – ability to recognise students' misconceptions, difficulties, and solution strategies
  - Instruction – teachers' knowledge of different representation and explanations of standard problems
    - Curricular level (indicator of cognitive activation)
    - Individual learning support (teachers providing adaptive explanations)
    - Effective classroom management

## Dunlosky et al. (2013)

Dunlosky et al. (2013) produce an extensive monograph that explores ten common learning practices. The selection of these practices is not meant to be exhaustive, but to cover a few that were widely identified as common, as well as a few that would be "easy to use." To this end, it is not meant to provide a complete framework of effective teaching. Instead, it reviews the strengths and shortcomings of each of these practices.

The ten techniques reviewed by the researchers include: elaborative interrogation; self-explanation; summarization; highlighting; keywork mnemonic; imagery use for text learning; rereading; practice testing; distributed practice; and interleaved practice.

For each of these ten, the authors describe key research that explains the technique and how it can be implemented. A notable strength of the article is how they also discuss the generalisability across different contexts—including student characteristics and learning indicators. They offer these reviews and critiques independently, rather than as a unified proposal to implement the entire set.

As a result, they offer different assessments of these ten techniques. Practice testing and distributed practice are deemed to be highly effective practices; elaborative interrogation, self-explanation, and interleaved practice are deemed moderately effective. The remaining five are classified with a low utility level. The researchers, however, do not argue that these techniques should be completely abandoned.

Instead, due to the limited contexts or criteria wherein they appear to be effective, coupled with a lack of rigorous evidence, the conclusion is that their role is limited.

Care should be taken to view these ten techniques as a guide for teachers. While the reviews (as the article reads more as multiple semi-independent reviews rather than one single one) of the literature are extensive, the ten topics discussed do not cover every classroom practice. Indeed, the purpose of the review was not to cover all classroom practices, but certain learning techniques. This exclusive focus on cognitive and learning sciences is certainly important, but does not claim to cover the complete range of an effective teacher's actions. Furthermore, this article is incredibly lengthy and laden with technical references and extensive citations. While this is certainly a strength of this source as an academic piece, it is not in a format that is accessible to most teachers. To Dunlosky's great credit, parallel versions have been created that communicate key findings in a medium and style that is accessible and useful to teachers.

The reviews effectively offer researchers some techniques that compose effective teaching, but ultimately stop short of suggesting all the practices effective teaching comprises.

## Praetorius et al. (2018)

Praetorius et al. (2018) present a framework for teaching quality that has been widely used in German-speaking countries and was originally developed in the context of maths education for the 1995 TIMSS-Video study. A three-dimensional model emerged from factor analysis of these instruments. The main framework consists of the three main dimensions, beneath which are 21 sub-dimensions. These sub-dimensions are derived from a set of classroom observation scales developed in Germany in the 1990s (Clausen, 2002; Gruehn, 2000). For each sub-dimension, Praetorius et al. give up to three example items to illustrate how it has been operationalised.

One feature of this model is that it contains nothing that is subject-specific: "the dimensions are conceptualised as being generic in nature, and thus as being applicable across school subjects" (p. 2).

The Three Basic Dimensions framework is derived from a theoretically-guided view of teaching and learning, as much as by direct empirical evidence. For example, its view of motivation comes from Deci and Ryan's (2008) self-determination theory, focusing on competence, autonomy and relatedness as the requirements for students to be motivated.

## *The Three Basic Dimensions Framework*

The three main dimensions are classroom management, student support and cognitive activation. Their components, as listed by Praetorius et al. (2018) are set out here:

***Classroom management***

- (Lack of) disruptions and discipline
- (Effective) time use/time on task
- Monitoring/'withitness'
- Clear rules and routines

***Student support***

- Support of competence experience
  - Differentiation and adaptive support
  - Pace of instruction
  - Constructive approach to errors
  - Factual, constructive feedback/appreciation
- Support of autonomy experience
  - Interestingness and relevance
  - Performance pressure and competition (negative indicator)
  - Individual choice options
- Support of social relatedness experience
  - Teacher → student
  - Student → teacher
  - Student → student

***Cognitive activation***

- Challenging tasks and questions
- Exploring and activating prior knowledge
- Exploration of the students' ways of thinking/elicitation of student thinking
- Receptive/transmissive understanding of learning of the teacher (negative indicator)
- Discursive and co-constructive learning
- Genetic-socratic teaching
- Supporting metacognition

## *Supporting evidence*

Praetorius et al. cite 39 research reports, based on 21 research studies/projects, in support of the framework. For each sub-dimension in the framework, the number of studies included and reported is shown in Table 1.

***Table 1: Number of studies included in each element of the three-dimensional model in Praetorius et al. (2018)***

| | | No. of cited studies that included this element |
|---|---|---|
| **Classroom management** | | |
| | (Lack of) disruptions and discipline | 17 |
| | (Effective) time use/time on task | 15 |
| | Monitoring/'withitness' | 6 |
| | Clear rules and routines | 5 |
| **Student support** | | |
| | Differentiation and adaptive support | 12 |
| | Pace of instruction | 6 |
| | Constructive approach to errors | 11 |
| | Factual, constructive feedback/appreciation | 4 |
| | Interestingness and relevance | 5 |
| | Performance pressure and competition (negative indicator) | 4 |
| | Individual choice options | 6 |
| | Teacher → student | 14 |
| | Student → teacher | 4 |
| | Student → student | 6 |
| **Cognitive activation** | | |
| | Challenging tasks and questions | 16 |
| | Exploring and activating prior knowledge | 7 |
| | Exploration of the students' ways of thinking/elicitation of student thinking | 8 |
| | Receptive/transmissive understanding of learning of the teacher (negative indicator) | 2 |
| | Discursive and co-constructive learning | 5 |
| | Genetic-socratic teaching | 3 |
| | Supporting metacognition | 2 |

We can see that no sub-dimension was included in all the studies and only a handful were in more than half.

Praetorius et al. also provide evidence of the predictive validity of the elements of the framework: the extent to which they predict learning gains and other outcomes. Correlations are reported at the level of the three basic dimensions (i.e., classroom management, student support, cognitive activation). These dimensions have been measured in a number of different modes: from classroom observation, from student surveys, from teacher self-report, from analysis of classroom artefacts. The outcomes used also vary across studies, ranging from before-and-after standardised curriculum assessments to self-reports of student enjoyment. For the before-and-after measures, the time gap between them also varies: from nine lessons, to one year.

If we limit our focus to studies where the outcome is gains on some kind of assessment of attainment, there are 25 level-2 (classroom) regression coefficients, ranging from -0.27 to 0.46, with median coefficients of 0.18, 0.12 and 0.17 for classroom management, student support and cognitive activation, respectively.[4]

Overall, predictive validity is quite low, and mixed. Even when the best measures are combined, collectively these measures of classroom quality do not explain much of the variation in student learning gains. The authors themselves note that "the results regarding the predictive validity of the Three Basic Dimensions are not convincing" (p. 16).

The authors also acknowledge that there may be other important characteristics of effective teaching not captured in the framework. Gaps might include generic and content-specific elements (p. 17).

All in all, this study offers a useful contribution to the Great Teaching Toolkit. It is grounded in strong empirical data and sound theory, and has been robustly tested in both research and practice. Its evidence base is purely correlational, though it does draw on a range of methods (e.g., observation, student survey), and while correlations are not large (0.1 – 0.2), they are consistent with other studies.

The work of Praetorius et al. makes the case for three broad dimensions in the Great Teaching Toolkit, which seems to represent a reasonable consensus across many other studies. However, the exact contribution of each of the sub-dimensions is less clear. They are certainly useful as exemplars and operationalised elements; whether they are all important for teacher development is unknown.

## Seidel and Shavelson (2008)

This meta-analysis makes a number of contributions to our knowledge about the characteristics of effective teaching.

First, as a systematic review and meta-analysis of studies between 1995 and 2004, it provides an important summary of the evidence from that period, though it also sets out to problematise the whole endeavour of meta-analysis. Evidence is synthesised from 112 publications, containing 1,357 estimates of the relationship between a range of instructional or classroom components and student outcomes, adjusted for background characteristics.

4 Taken from Table 4 in Praetorius et al., 2018.

Second, it gives a contextualised summary of and comparison with the results from two previous systematic reviews, by Fraser et al. (1987) and Scheerens and Bosker (1997). While the latter study reported an overall correlation of 0.25 between observable teaching quality and outcomes, Seidel and Shavelson, a decade later (and using a similar theoretical model), report an average of 0.02, with no individual component achieving a correlation higher than 0.04. Their main explanation for this difference is not that the strength of relationship has reduced, but that the later review used better controls[5] and weighting of component study estimates – both of which are standard practice in high-quality meta-analysis today.

Third, it presents, as an alternative to the traditional 'process-product' approach, a more theoretically-guided cognitive model of teaching and learning as a conceptual framework for the meta-analysis. This model focuses on the functions, purpose and context of different classroom components. These are identified as a set of contextual factors and a set of teacher practices, taken from a model outlined by Bolhuis (2003). The contextual factors are: (1) knowledge domain (the subject or curriculum being taught); (2) time for learning; (3) organisation of learning (classroom management); and (4) social context (social learning climate). The teacher practices are: (5) goal-setting/orientation (e.g., clarifying goals, teaching in a clear and structured way, activating student pre-knowledge); (6) execution of learning activities (providing opportunities for processing information, such as modelling, problem-solving and inquiry); (7) evaluation (assessment); and (8) regulation and monitoring (giving feedback and promoting meta-cognitive strategies). This framework is used to classify the effects from different studies and more variation is found, particularly when the 'effects' are separated according to the method of measuring the instructional and classroom components (whether by teacher survey, student survey or observation/video analysis) or split by type of design (experimental/quasi-experimental vs correlational).

Seidel and Shavelson's main substantive finding is that "we found the largest teaching effects for domain-specific components of teaching—teaching components most proximal to executive learning processes". However, it is not very clear exactly what kinds of teacher practices were classified under that heading, beyond the examples cited of "activities such as mathematical problem-solving, scientific inquiry, or specific reading and writing strategies". "Organisation of learning" (i.e., classroom management) was also found to have a consistently strong relationship with learning outcomes.

## Creemers and Kyriakides (2006; 2011): Dynamic Model

Creemers and Kyriakides (2006; 2011), and their Dynamic Model, come from the tradition of 'Educational Effectiveness Research' (EER, a blending of previously separate traditions of School Effectiveness Research with studies of classroom practices and teacher effectiveness, some of which go back to the 1960s; Creemers & Kyriakides, 2015; Creemers et al., 2013). This research tradition is characterised by the use of assessments of student learning, often limited to assessments of maths and reading in standardised tests, and statistical models (often multi-level regression models) to adjust for a range of covariates, interpreting the unexplained variation in outcomes as the 'effect' of the school or teacher.

The Dynamic Model takes this work forward by recognising greater complexity in a number of ways. First is the insight that understanding the relationships between educational inputs (resources, behaviours,

5 In particular, they did not use correlations between an observed practice and raw student outcomes unless some adjustment for prior covariates was made (e.g., prior attainment or SES).

policies, etc.) and outputs (e.g., learning) requires a range of different types of theory. In particular, if we want to improve as well as understand, this theory is even more important – especially, good theories of teaching and learning. Related to this, the authors note that they draw on a wide range of different perspectives on pedagogy, including elements that would generally be associated with both direct instruction approaches and constructivism (Creemers et al., 2013).

The weaving together of these two strands is a key element of the Dynamic Model. On the one hand, Creemers et al. (2013) characterise the 'competency-based approach' as setting out to list explicit strategies and competences, drawing on the 'process-product' tradition of work by researchers such as Good and Brophy (1984) and Rosenshine (1976), who observed consistent empirical associations between certain observable teacher behaviours and student achievement. This leads to a view of effective teaching as skills-based and discrete skills, such as classroom management, clear and concise exposition of ideas, and using questioning, modelling, reviewing and feedback. Theories such as Carroll's (1963) model of learning and cognitive load theory (Sweller et al., 1998) are often invoked in this tradition and approaches such as mastery learning (Bloom, 1976) or direct instruction (Rosenshine, 1987) can be seen as practical, packaged instantiations of it.

On the other hand, Creemers et al. (2013) contrast this with a more holistic approach to understanding teaching quality. This tradition draws on the ideas of 'reflection in action' derived from Dewey (1933), Stenhouse (1975) and Schon (1983), that stress the need for teachers' own critical reflection on their practice, and the work of developmental psychologists such as Piaget (1972), who emphasise the need for learners (including teachers learning their practice) to actively construct meaning from experience. Becoming a great teacher requires more than just practising isolated techniques: each teacher must understand their own context, reflect on their practice, and, through action research and teacher inquiry, find their own solutions to the problems it presents. Teaching cannot be reduced to a mechanistic, technicist project: it is just too complex to have a single right way that can be simply described and applied universally. Professional development should emphasise critical reflection, inquiry, agency and moral purpose.

While these two approaches are often seen as incompatible paradigms – and even in opposition, dividing educators between 'traditional' and 'progressive' camps – for Creemers et al. (2013), they each offer part of the story. Research evidence is clear that certain practices and techniques are powerful determinants of student learning: teachers should know about these competences and should be supported in learning them. However, a great teacher is more than just a set of isolated competences: teachers must understand the underlying theory and processes, as well as knowing their own context, so that they can make judicious selections and adaptations, and implement these practices faithfully and effectively. Moreover, according to the evidence from the Dynamic Model, the emphasis of teachers' professional learning should reflect the stage of their development: for less effective (typically early-career) teachers, developing the basic skills is the priority; once those skills are mastered, they have something to reflect with and can apply that reflection to the more complex challenges of refining, orchestrating and adapting their practices.

The second insight of the Dynamic Model is that there are a range of outcomes of interest, reflecting overlapping educational aims – for example, cognitive, psychomotor, metacognitive and affective – not just basic skills. Moreover, some of the validation studies have evaluated the equity of educational outcomes as well as their overall levels. Third, the factors influencing effectiveness operate at multiple levels (student, classroom, school and system). Fourth, these factors may also interact, either within a level or across levels. Some characteristics of effectiveness may act more like catalysts or barriers, in

combination with other factors: the effect of each factor may depend on other factors. Fifth, and related to the previous point, some factors may be expected to have non-linear relationships with outcomes. For example, it could be the case that more of a particular factor is associated with better outcomes up to a point, but beyond that the relationship plateaus or even reverses.

The model specifies what these factors are, based on empirical evidence and tested theory. Where other models typically focus on measuring the amount of a factor, the Dynamic Model recognises that the factors differ qualitatively as well as quantitatively. For each factor, as well as its frequency (how much or how often it occurs), the model captures its focus (its function, specificity and purpose), stage (the duration and timing of a policy or practice), quality (whether it is done well, in ways that are aligned with best evidence) and differentiation (the extent to which its implementation is adapted to individual student/classroom/school context and needs).

At the student level, the Dynamic Model includes the following (Creemers & Kyriakides, 2011, p. 29):

- Socio-cultural and economic factors, such as socioeconomic status, ethnicity and gender, their interactions and compositional effects
- Psychological characteristics: aptitude, motivation, expectations, personality and thinking style[6]
- Opportunity to learn: the time made available and focused on learning activities aligned with the intended curriculum
- Time on task: time actually spent engaged in learning, limited by individual student interest, motivation, expectations, attention, self-regulation and quality of focus

Classroom factors relate to the behaviour of the teacher and 'refer to observable instructional behaviours of teachers in the classroom rather than on factors that may explain such behaviours (e.g., teacher beliefs and knowledge and interpersonal competences)'. There are eight factors in the model:

6 The justification for 'thinking style' draws on the work of Sternberg (e.g., 1988) that tries to account for differences in performance and 'intelligence' in terms of intellectual styles of mental self-government.

***Table 2: The dynamic model of educational effectiveness (Creemers & Kyriakides, 2011, p. 35)***

| | |
|---|---|
| (1) Orientation | (a) Providing the objectives for which a specific task/lesson/series of lessons takes place |
| | (b) Challenging students to identify the reason why an activity is taking place in the lesson |
| (2) Structuring | (a) Beginning with overviews and/or reviews of objectives |
| | (b) Outlining the content to be covered and signalling transitions between lesson parts |
| | (c) Drawing attention to and reviewing main ideas |
| (3) Questioning | (a) Raising different types of questions (i.e., process and product) at appropriate difficulty level |
| | (b) Giving time for students to respond |
| | (c) Dealing with student responses |
| (4) Teaching modelling<br>. | (a) Encouraging students to use problem-solving strategies presented by the teacher or other classmates |
| | (b) Inviting students to develop strategies |
| | (c) Promoting the idea of modelling |
| (5) Application | (a) Using seatwork or small-group tasks in order to provide needed practice and application opportunities |
| | (b) Using application tasks as starting points for the next step of teaching and learning |
| (6) The classroom as a learning environment | (a) Establishing on-task behaviour through the interactions they promote (i.e., teacher–student and student–student interactions) |
| | (b) Dealing with classroom disorder and student competition through establishing rules, persuading students to respect them and using the rules |
| (7) Management of time | (a) Organising the classroom environment |
| | (b) Maximising engagement rates |
| (8) Assessment | (a) Using appropriate techniques to collect data on student knowledge and skills |
| | (b) Analysing data in order to identify student needs and report the results to students and parents |
| | (c) Teachers evaluating their own practices |

The Dynamic Model has a number of very specific strengths.

First is the fact that the model is derived from a large body of previous evidence about the effectiveness of schools and teachers.

Second, the Dynamic Model has been subjected to a fair level of rigorous testing of its claims and predictions, albeit mainly by the authors themselves. For example, a 2013 meta-analysis by Kyriakides et al. broadly supports this framework; Creemers et al. (2012) also provide a review of evidence from several studies.

Third, the model is accompanied by an extensive suite of instruments to measure its various components. Creemers and Kyriakides (2011) provide details of these instruments that include student questionnaires, classroom observation schedules (high and low inference), and teacher questionnaires.

A fourth strength is that the implications of the Dynamic Model for school and teacher improvement have been developed and tested in practice. In the study reported by Creemers et al. (2013), 130 primary school teachers in Cyprus were randomly allocated to either a development programme based on the Dynamic Model (the 'Dynamic Integrated Approach' – DIA) or an alternative 'holistic approach', based on supporting teachers' critical reflection on their pedagogy. Teachers were assessed at the start as fitting one of five developmental stages, and randomisation was 'blocked' within each stage (i.e., each stage had equal numbers in each treatment arm). A brief description of the characteristics of the five stages is given in Table 3. Both arms had eight PD sessions, approximately once a month over a school year, in which they developed an individual action plan and received support from peers as well as from expert outsiders, who also observed and gave feedback.

In the DIA treatment arm, teachers were grouped according to their developmental stage and given materials and training specific to their stage. In developing their action plans, they were encouraged to focus on stage-appropriate, research-backed approaches and their progress was supported with ongoing provision of targeted reading materials and tasks. By contrast, teachers in the 'holistic approach' arm were free to choose their own goals and activities, supported by peer-group and expert-moderated discussion and critical reflection.

A number of outcomes were recorded. In terms of observational assessments of teaching quality, about a third of the DIA teachers moved up a stage, compared with none of those in the holistic arm. All teachers in the DIA group improved their quality scores at least to some extent and the mean change was equivalent to about 0.6 SD (compared with 0 for the 'holistic' group). Neither group exhibited any change in their attitudes or perceptions. Assessments of student learning showed an effect size of 0.24 in favour of those in the DIA arm, with larger effects (an additional 0.14) for those teachers who progressed a whole stage. A one-year follow-up of both groups found that the change (or lack of change) in teaching quality had been maintained (Creemers et al., 2013, p. 218).

Creemers et al. (2013) interpret these results as showing that "reflection is more effective when teachers' priorities for improvement are taken into account and when they are encouraged to develop action plans which address their professional needs; these were identified through a relevant empirical investigation". They note that some of the teachers in the holistic arm selected areas to work on that were well above their stage, for example, trying to differentiate their instruction when their classroom management skills were weak: "their attempts to incorporate this into their teaching were not successful" (p. 178).

***Table 3: Five stages of teaching skills (based on Creemers et al., 2013, pp. 163, 179)***

| Stage | Teaching skills | Focus for development |
|---|---|---|
| 1. Basic elements of direct teaching | • Management of time<br>• Use of structuring, application, assessment and questioning<br>• Some attention to teacher-student relations | *Maximising opportunity to learn*<br>• Lesson structuring (sequencing, connecting to previous, key points explained)<br>• Use of application activities/exercises (practising application and implementation of knowledge/procedures, giving individual feedback and asking questions)<br>• Questioning and providing feedback (asking many questions of all students, giving them time to reflect) |
| 2. Putting aspects of quality in direct teaching and touching on active teaching | • More sophisticated use of structuring, application, questioning (i.e., appropriate for stage and purpose in relation to context)<br>• Some attention to student relations | *Matching lesson activities to purpose and context*<br>• Timing of the application tasks (judging when to use and what knowledge, skills, applications and links to previous learning to include)<br>• Quality of the lesson structuring (effective connections with and review of previous learning, highlighting key points) |
| 3. Acquiring quality in active/direct teaching | • More sophisticated attention to teacher-student and student-student relations (appropriate timing and purpose)<br>• More sophisticated use of assessment, orientation, feedback, questioning and structuring (in relation to timing and purpose, but also quality)<br>• Use of teaching modelling | *Developing the classroom learning environment for constructing knowledge*<br>• Orientation to learning aims (ensuring students understand learning goals)<br>• Development of the classroom as a learning environment (regular, high-quality, learning-focused interactions between teacher and students, and between students; students are encouraged to express their views or explore different solutions, but also challenged to justify them) |
| 4. Differentiation of teaching | • Appropriately differentiated and high-quality use of structuring, time management, questioning, application, assessment, teaching modelling and orientation | *Differentiating appropriately*<br>• Differentiation of teaching (teachers adapting their approach to student characteristics, readiness and needs, in their questioning, feedback, assignment of application tasks and follow-up)<br>• Orientation of students to the learning goals (strategies for engaging students in identifying learning objectives: 'why are we doing this?') |

| | |
|---|---|
| 5. Achieving quality and differentiation in teaching using different approaches | • Use of differentiated and high-quality practices in relation to teacher-student and student relations<br>• Appropriately differentiated, appropriate and high-quality use of orientation and teaching modelling |

In a partial replication of this study, Creemers et al. (2013) compared groups of teachers in a four-arm randomised trial with a 2x2 factorial design. As before, two of the groups used the DIA approach and two the holistic approach, but this time one of each was supported by external experts (as above) while the other was supported by colleagues within their own school, using a programme designed by the researchers. Results were very similar to the previous study, with gains in observed teaching competence and student attainment for the DIA group, but none for the 'holistic' approach. Interestingly, there was no difference between teachers supported directly by external experts and those supported by their colleagues.

A further evaluation of the DIA (Creemers et al., 2013) compared training a group of teachers in their knowledge and use of assessment. Again, the DIA group were assessed (this time on their use of assessment, using a self-report questionnaire), divided into groups based on their stage and given learning materials and support that were targeted at their level. In this study, the comparison was with a randomly equivalent group allocated to be given skills-based training[7] in assessment, but where the training was the same for all, irrespective of their existing level of practice; there was also a third, no-treatment, control arm. Assessments of their use of assessment (from teacher self-report) improved for both the active groups, but not for the controls. Gains for the 'targeted' (DIA) group were about twice the size of those for the generic skills-based training. There were also gains in student achievement: for teachers judged to be at stage 1 (of 4) in their use of assessment, both the active groups saw improved attainment equivalent to an effect size of about 0.1. For teachers at the higher stages, only the DIA group improved (ES=0.17).

The main takeaway from these professional development studies using the Dynamic Model seems to be that we can usefully identify stages of teacher effectiveness. Despite the diversity of different elements of classroom quality, they seem to cluster into levels. The impact of professional development on student learning seems to be greatest when it is targeted at developing practices and skills that take each teacher from their current stage to the next.

## Scheerens et al. (2007) meta-analysis

Scheerens et al. (2007) present a review, conceptualisation and meta-analysis of both school-level and classroom-level factors related to student outcomes. They identify 46 factors at the classroom level and categorise them into 15 teaching dimensions. The mean correlation[8] and number of effects are shown in Table 4.

7 Derived from the competency-based approach, described above.

8 Technically, these are Fisher-Z scores. For the range of values here, they are within 5% of the corresponding correlations.

*Table 4: Effect sizes (Fisher-Z) of classroom factors from Scheerens et al. (2007)*

| Classroom factor | Mean Effect | No of effects | |
|---|---|---|---|
| **1** | **Learning Time** | **.095** | |
| 1.1 | time on task | .125 | 86 |
| 1.2 | opportunity to learn | .118 | 32 |
| 1.3 | homework | .041 | 51 |
| 1.4 | mastery learning | .047 | 4 |
| **2** | **Classroom organisation** | **.075** | |
| 2.1 | classroom management | .088 | 36 |
| 2.2 | discipline | .070 | 20 |
| 2.3 | control | .018 | 17 |
| **3** | **Learning environment** | **.129** | |
| 3.1 | classroom climate | .125 | 107 |
| 3.2 | no achievement pressure | .151 | 29 |
| 3.3 | mastery orientation | -.005 | 2 |
| 3.4 | no performance orientation | .120 | 2 |
| **4** | **Clear and structured** | **.126** | |
| 4.1 | structured/direct teaching | .107 | 76 |
| 4.2 | goal-directed/clear | .222 | 36 |
| 4.3 | teacher demonstration | .014 | 17 |
| 4.4 | teaching basic skills | .073 | 17 |
| **5** | **Activating** | **.123** | |
| 5.1 | cooperative | .204 | 49 |
| 5.2 | situated/discovery | .155 | 3 |
| 5.3 | peer tutoring | .218 | 53 |
| 5.4 | student work | .059 | 36 |
| 5.5 | individual work | -.009 | 39 |
| 5.6 | student discussions | .043 | 8 |
| **6** | **Learning strategies** | **.213** | |
| 6.2 | meta-cognitive | .244 | 35 |
| 6.3 | scientific inquiry | .197 | 32 |
| 6.5 | organising methods | .000 | 2 |
| 6.7 | reading/writing | .210 | 34 |
| **7** | **Challenge** | **.130** | |
| 7.1 | cognitive activation/understanding orientation | .182 | 67 |
| 7.2 | active student engagement | .042 | 63 |
| 7.3 | authentic contexts/relevance | .160 | 47 |
| 7.4 | language level | .029 | 7 |
| 7.5 | representation formats | .385 | 4 |
| **8** | **Support** | **.108** | |
| 8.1 | quality of interactions/teacher support | .108 | 73 |

| | | | |
|---|---|---|---|
| **9** | **Feedback** | **.056** | |
| 9.1 | feedback/frame of reference/monitoring | .056 | 106 |
| **10** | **Evaluation** | **.086** | |
| 10.1 | assessments/tests | .086 | 46 |
| **11** | **Teacher characteristics** | **.146** | |
| 11.1 | high expectations | .124 | 22 |
| 11.2 | constructivist beliefs about learning | .354 | 4 |
| **12** | **Adaptive Teaching** | **.066** | |
| 12.1 | various teaching methods | .124 | 2 |
| 12.2 | adaptive teaching | .036 | 27 |
| 12.3 | open tasks/choice | .090 | 4 |
| 12.4 | student prerequisites | .178 | 7 |
| **13** | **Practice** | **-.080** | |
| 13.1 | drill/repetition | -.078 | 17 |
| 13.2 | application | -.057 | 19 |
| **14** | **Material** | **.015** | |
| 14.1 | textbooks | .039 | 6 |
| 14.2 | media | .012 | 27 |
| **15** | **Integrative approaches** | **.089** | |
| 15.1 | constructivist | .039 | 52 |
| 15.2 | inductive | -.197 | 5 |
| 15.3 | concept-oriented | .257 | 33 |

Scheerens et al. note some interesting results. The largest individual factors (e.g., representation formats and constructivist beliefs about learning) come from quite small numbers of studies so should be interpreted cautiously. Among those with more replications, teaching meta-cognitive strategies, peer tutoring, cooperative learning, and instruction that is clearly goal-directed have relatively high coefficients (above 0.2). Similarly high coefficients are also found for subject-specific learning strategies, like scientific inquiry and reading and writing. A number of these larger effects are for factors associated with 'constructivist' approaches, compared with negative effects for practice (drill/repetition, application) and for performance pressure (i.e., positive correlations for no achievement pressure and no performance orientation).

A comparison that groups 'constructivist-oriented' approaches (cooperative, situated/discovery, peer tutoring, student work, individual work, student discussions, meta-cognitive, scientific inquiry, organising methods, reading/writing, cognitive activation/understanding orientation, active student engagement, authentic contexts/relevance, constructivist beliefs about learning, constructivist, inductive, concept-oriented) against 'structured/direct/mastery' approaches (mastery learning, **mastery** orientation, structured/direct teaching, goal-directed/clear, teacher demonstration, teaching basic skills, drill/repetition, application) narrowly favours the former.

## Danielson (2007): Framework for Teaching

The Framework for Teaching was created by Charlotte Danielson and particularly grew in popularity during the 2010s. It presents four domains, divided into a total of 22 components. They are: planning and preparation, classroom environment, instruction, and professional responsibilities. In each of these domains, teachers can be labelled as unsatisfactory, basic, proficient, or distinguished.

The *planning and preparation* domain refers not just to the design and preparation of teaching, but also the content knowledge that the teaching entails. The components of this domain also include the implementation of assessments, as well as instruction that is aligned to the curriculum. The second domain, *classroom environment*, refers to both the physical space and the classroom as a social space. This then includes appropriate student behaviour as a critical component. *Instruction* refers to 'engaged learning', which covers students actively engaging with materials at a high level. There are also elements of student metacognition in this domain, with students understanding learning goals. The final domain, *professional responsibilities*, include additional responsibilities placed on teachers that contribute to a school's success. These range from record-keeping, communicating with families, and professional development.

The Framework for Teaching includes a range of behaviours and expectations that extend beyond a definition of effective *teaching*; it could be said they aim to describe a good *teacher*. That is, the professional aspects that extend beyond formal learning (e.g., maintaining accurate records) bear as much weight as instructional ones (e.g., engaging students in learning).

The framework is widely accessed by school leaders; in some cases, it is used as the basis for teacher evaluations. While resources for teachers and leaders can be readily obtained online, greater detail can be found in Danielson's book. The domains and components were developed by both 'practice wisdom' and underlying research; this research, however, is not as readily available as the framework itself.

## Early Career Framework (2019)

The Department for Education in England consulted widely with researchers, as well as the wider education sector, to develop the Early Career Framework. In support of this, the Education Endowment Foundation acted as an independent reviewer to ensure the framework is robust and evidence-based. Explicitly not intended to be an assessment rubric, the framework is meant to support and "underpin" professional development for early career teachers.

The framework has eight dimensions. Each enumerates a series of statements of what teachers should learn related to the dimension, as well as the behaviours they should exhibit. The framework drew the "learn that..." statements from the best available evidence; the "learn how to..." statements were additionally collected from expert practitioners.

- **High expectations**
  Learn how to...
  - Communicate a belief in the academic potential of all pupils
  - Demonstrate consistently high behaviour expectations
- **How pupils learn**
  Learn how to...
  - Avoid overloading working memory
  - Build on pupils' prior knowledge

    - Increase likelihood of material being retained
- **Subject and curriculum**
  Learn how to...
    - Deliver a carefully sequenced and coherent curriculum
    - Support pupils to build increasingly complex mental models
    - Develop fluency
    - Help pupils apply knowledge and skills to other contexts
    - Develop pupils' literacy
- **Classroom practice**
  Learn how to...
    - Plan effective lessons
    - Make good use of expositions
    - Model effectively
    - Stimulate pupil thinking and check for understanding
- **Adaptive teaching**
  Learn how to...
    - Develop an understanding of different pupil needs
    - Provide opportunity for all pupils to experience success
    - Meet individual needs without creating unnecessary workload
    - Group pupils effectively
- **Assessment**
  Learn how to...
    - Avoid common assessment pitfalls
    - Check prior knowledge and understanding during lessons
    - Provide high-quality feedback
    - Make marking manageable and effective
- **Managing behaviour**
  Learn how to...
    - Develop a positive, predictable and safe environment for pupils
    - Establish effective routines and expectations
    - Build trusting relationships
    - Motivate pupils
- **Professional behaviours**
  Learn how to...
    - Develop as a professional
    - Build effective working relationships
    - Manage workload and wellbeing

The framework includes both classroom- and learning-focused behaviours (with significant focus on learning and cognitive actions), as well as those of a teacher. It is responsive to current trends and needs in education in England, with workload appearing in multiple capacities. Similarly, understanding how pupils learn is the focus of a whole dimension on its own. Naturally, a strength of the framework is the inclusion of extensive research and resources, each linked to a corresponding dimension. It therefore effectively lays out a series of focuses for early career teachers for their own professional development.

# References and further reading


Adams, G., & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISTAR.* Educational Achievement Systems.

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045), 1034–1037. https://doi.org/10.1126/science.1207998

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. *Journal of Teacher Education, 59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development, 67*(3), 1206–1222. https://doi.org/10.1111/j.1467-8624.1996.tb01791.x

Baumert, J., & Kunter, M. (2013). The COACTIV Model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV Project* (pp. 25–48). Springer. https://doi.org/10.1007/978-1-4614-5149-5_2

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Bennett, T. (2017). *Creating a culture: How school leaders can optimise behaviour*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/602487/Tom_Bennett_Independent_Review_of_Behaviour_in_Schools.pdf

Bjork, E. L., & Bjork, R. A. (2009). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. R. Pomerantz (Eds.), *Psychology and the real world: essays illustrating fundamental contributions to society* (pp. 55–64). Worth Publishers.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education, 56*, 35–46. https://doi.org/10.1016/j.tate.2016.02.003

Bloom, B. S. (1976). Human characteristics and school learning. In *Human characteristics and school learning*. McGraw-Hill.

Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: A multidimensional perspective. *Learning and Instruction, 13*(3), 327–347. https://doi.

org/10.1016/S0959-4752(02)00008-7

Booth, J. L., McGinn, K. M., Barbieri, C., Begolli, K. N., Chang, B., Miller-Cotto, D., Young, L. K., & Davenport, J. L. (2017). Evidence for cognitive science principles that impact learning in mathematics. In D. C. Geary, D. B. Berch, R. J. Ochsendorf, & K. M. B. T.-A. of C. A. S. and H.-O. M. C. Koepke (Eds.), *Acquisition of complex Arithmetic skills and higher-order mathematics concepts* (pp. 297–325). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-805086-6.00013-8

Bowlby, J. (1969). Attachment and loss: Attachment (Vol. 1). In *Attachment* (Vol. 1). Basic Books. https://doi.org/978/0712674713

Braasch, J. L. G., Goldman, S. R., & Wiley, J. (2013). The influences of text and reader characteristics on learning from refutations in science texts. *Journal of Educational Psychology, 105*(3), 561–578. https://doi.org/10.1037/a0032627

Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of variation and prior knowledge on sbstract concept learning. *Cognition and Instruction, 33*(3), 226–256. https://doi.org/10.1080/07370008.2015.1067215

Caldarella, P., Larsen, R. A. A., Williams, L., Downs, K. R., Wills, H. P., & Wehby, J. H. (2020). Effects of teachers' praise-to-reprimand ratios on elementary students' on-task behaviour. *Educational Psychology*, 1–17. https://doi.org/10.1080/01443410.2020.1711872

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*(8), 723–733.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

CESE (Centre for Education Statistics and Evaluation). (2017). *Cognitive load theory: Research that teachers really need to understand Centre for Education Statistics and Evaluation*. https://www.cese.nsw.gov.au/images/stories/PDF/cognitive-load-theory-VR_AA3.pdf

Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive? : empirische Analysen zur Übereinstimmung, Konstrukt- und Kritieriumsvalidität*. Waxmann.

Coe, R. (2014, January 9). *Classroom observation: It's harder than you think*. CEM Blog. http://www.cem.org/blog/414/

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. Sutton Trust.

Coe, R. (1998). Can feedback improve teaching? A review of the social science literature with a view to identifying the conditions under which giving feedback to teachers will result in improved performance. *Research Papers in Education, 13*(1), 43–66. https://doi.org/10.1080/0267152980130104

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. In *School Effectiveness and School Improvement* (Vol. 17, Issue 3, pp. 347–366). Routledge. https://doi.org/10.1080/09243450600697242

Creemers, B. P. M., & Kyriakides, L. (2011). *Improving quality in education: Dynamic approaches to school improvement*. Routledge. https://doi.org/10.4324/9780203817537

Creemers, B., & Kyriakides, L. (2015). Process-product research: A cornerstone in educational effectiveness research. *The Journal of Classroom Interaction, 50*(2), 107–119. http://www.jstor.org/stable/44735492

Creemers, B., Kyriakides, L., & Antoniou, P. (2013). *Teacher professional development for improving quality of teaching*. Springer. https://doi.org/10.1007/978-94-007-5207-8

Csikszentmihalyi, M., & Schneider, B. (2000). Becoming adult: How teenagers prepare for the world of work. In *Becoming adult: How teenagers prepare for the world of work*. (pp. xx, 289–xx, 289). Basic Books.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives, 8*(1), 1. https://doi.org/10.14507/epaa.v8n1.2000

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology, 49*(3), 182–185. https://doi.org/10.1037/a0012801

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 53, Issue C, pp. 63–147). Academic Press. https://doi.org/10.1016/S0079-7421(10)53003-2

Department for Education. (2019). *Early Career Framework*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/773705/Early-Career_Framework.pdf

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process* (2nd ed.). D.C. Heath and Co.

Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review, 11, 1–26*. https://doi.org/10.1016/j.edurev.2013.11.002

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

Education Endowment Foundation (EEF). (2019). *Improving behaviour in schools*.

Education Endowment Foundation (EEF). (2018). *Metacognition and self-regulated learning: Guidance report*.

Education Endowment Foundation (EEF). (2020). *Improving mathematics in the Early Years and Key*

*Stage 1*.

Ericsson, K. A. (2009). *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (K. A. Ericsson (ed.)). Cambridge University Press. https://doi.org/10.1017/CBO9780511609817

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*(2), 147–252. https://doi.org/10.1016/0883-0355(87)90035-8

Good, T. L., & Brophy, J. E. (1984). *Looking in classrooms*. Harper & Row.

Gruehn, S. (2000). *Unterricht und schulisches Lernen : Schüler als Quellen der Unterrichtsbeschreibung*. Waxmann.

Guay, F., Ratelle, C. F., & Chanal, J. (2008). Optimal learning in optimal contexts: The role of self-determination in education. *Canadian Psychology/Psychologie Canadienne, 49*(3), 233–240. https://doi.org/10.1037/a0012758

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of metacognition in education*. Routledge/Taylor & Francis Group.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development, 85*(3), 1257–1274. https://doi.org/10.1111/cdev.12184

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Hattie, J., & Timperley, H. (2007). The Power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hill, H. C., & Charalambous, C. Y. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Lessons learned and open issues. *Journal of Curriculum Studies, 44*(4), 559–576. https://doi.org/10.1080/00220272.2012.716978

Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal, 55*(5), 1076–1112. https://doi.org/10.3102/0002831218769614

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. https://doi.org/10.3102/00028312042002371

Hogarth, R. M. (2001). *Educating intuition*. Chicago.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science, 24*(5), 379–385. https://doi.org/10.1177/0963721415591878

Jones, S. M., & Doolittle, E. J. (2017). Social and emotional learning: Introducing the issue. *Future of Children, 27*(1), 3–12. https://doi.org/10.1353/foc.2017.0000

Kaiser, G., & König, J. (2019). Competence measurement in (mathematics) teacher education and beyond: Implications for policy. *Higher Education Policy, 32*(4), 597–615. https://doi.

org/10.1057/s41307-019-00139-z

Kennedy, M. (2016). Parsing the Practice of Teaching. *Journal of Teacher Education, 67*(1), 6–17. https://doi.org/10.1177/0022487115614617

Kern, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools, 44*(1), 65–75. https://doi.org/10.1002/pits.20206

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254.

Ko, J., Sammons, P., & Bakkum, L. (2013). *Effective teaching: A review of research and evidence*. CfBT Education Trust. https://eric.ed.gov/?id=ED546794

Kounin, J. S. (1977). *Discipline and group management in classrooms*. R.E. Krieger Pub. Co.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143–152. https://doi.org/10.1016/j.tate.2013.07.010

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465–491. https://doi.org/10.3102/00028312032003465

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

London, M. (2003). *Job feedback: Giving, seeking, and using feedback for performance improvement*. Lawrence Erlbaum Associates.

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis, 41*(3), 260–293. https://doi.org/10.3102/0162373719849044

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics, 99*(2), 486–496. https://doi.org/10.1016/j.jdeveco.2012.06.002

Meyer, J. H. F., & Land, R. (2005). Threshold concepts and troublesome knowledge (2): Epistemological considerations and a conceptual framework for teaching and learning. *Higher Education, 49*(3), 373–388. https://doi.org/10.1007/s10734-004-6779-5

Moore, D., Benham-Clarke, S., Kenchington, R., Boyle, C., Ford, T., Hayes, R., & Rogers, M. (2019). *Improving behaviour in schools: Evidence review*. https://educationendowmentfoundation.org.uk/public/files/Improving_B

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*(1), 33–52. https://doi.

org/10.1037/0022-3514.75.1.33

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art - teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451

Muijs, D., Reynolds, D., Sammons, P., Kyriakides, L., Creemers, B. P. M., & Teddlie, C. (2018). Assessing individual lessons using a generic teacher observation instrument: How useful is the International System for Teacher Observation and Feedback (ISTOF)? *ZDM - Mathematics Education, 50*(3), 395–406. https://doi.org/10.1007/s11858-018-0921-9

Nuthall, G. (2007). *The hidden lives of learners*. NZCER Press.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development*. https://doi.org/10.1159/000271225

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 365–386). Springer. https://doi.org/10.1007/978-1-4614-2018-7_17

Praetorius, A. K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM - Mathematics Education, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Rosenshine, B. (1976). Classroom instruction. In N. L. Gage (Ed.), *The psychology of teaching methods* (pp. 335–371). University of Chicago Press.

Rosenshine, B. (1987). Direct instruction. In M. J. Dunkin (Ed.), *International ecyclopedia of teaching and teacher evaluation* (pp. 257–262). Pergamon Press.

Rosenshine, B. (2010). Principles of instruction. *Educational Practices Series, 21*, 109–125. https://doi.org/10.1007/978-94-007-2669-7_7

Rosenshine, B., & Stevens, R. (1986). Teaching functions. *Handbook of Research on Teaching, January 1986*, 376–391.

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal, 50*(5), 1020–1049. https://doi.org/10.3102/0002831213477680

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Pergamon Press.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*.

Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Shimamura, A. (2018). *MARGE. A whole-brain learning approach for students and teachers*. https://shimamurapubs.files.wordpress.com/2018/09/marge_shimamura.pdf

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Smith, T. W., Baker, W. K., Hattie, J., & Bond, L. (2008). A validity study of the certification system of the National Board for Professional Teaching Standards. In R. E. Stake, S. Kushner, L. Ingvarson, & J. Hattie (Eds.), *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards* (Vol. 11, pp. 345–378). Emerald Group Publishing Limited. https://doi.org/10.1016/S1474-7863(07)11012-7

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176–199. https://doi.org/10.1177/1745691615569000

Stenhouse, L. (1975). *An introduction to curriculum research and development*. Pearson Education.

Sternberg, R. J. (1988). Mental self-government: A theory of intellectual styles and their development. *Human Development, 31*(4), 197–224. https://doi.org/10.1159/000275810

Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research, 88*(4), 479–507. https://doi.org/10.3102/0034654317751919

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. In *Educational Psychology Review* (Vol. 31, Issue 2, pp. 261–292). Springer New York LLC. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development*. 344. http://educationcounts.edcentre.govt.nz/goto/BES

van de Grift, W. J. C. M., Chun, S., Maulana, R., Lee, O., & Helms-Lorenz, M. (2017). Measuring teaching quality and student engagement in South Korea and The Netherlands. *School Effectiveness and School Improvement, 28*(3), 337–349. https://doi.org/10.1080/09243453.2016.1263215

van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology, 20*(3), 343–352. https://doi.org/10.1002/acp.1250

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*(2), 246–260. https://doi.org/10.1037/0022-3514.87.2.246

Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar (Trans.)). MIT Press.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122. https://doi.org/10.3102/00346543073001089

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Weinstein, Y., Sumeracki, M., & Caviglioli, O. (2018). *Understanding how we learn: A visual guide.* Routledge.

Wiliam, D. (2018). *Creating the schools our children need: Why what we're doing now won't help much (and what we can do instead)*. Learning Scineces International.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). Routledge. https://doi.org/https://doi.org/10.4324/9780203874851

Willingham, D. T. (2019, July 14). *The high price of multitasking*. The New York Times. https://www.nytimes.com/2019/07/14/opinion/multitasking-brain.html

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81*(2), 267–301. https://doi.org/10.3102/0034654311405999