How might we take this forward?

This final section of our review pulls together the implications of the research evidence we have presented and proposes a framework for conceptualising teaching quality. We then make some recommendations for practitioners about how these ideas could be used to promote better teaching.

Overview of the evidence

Evidence about effective pedagogy

In Section 2 (p9) we identified a selection of teaching approaches, skills and knowledge that have been shown to be related to enhanced student outcomes. The evidence here is often weak or equivocal, and it is easy to select from it to make claims that fit preconceptions. The effective practices themselves are often quite loosely described, leaving room for interpretation about whether what one has observed is in fact an example of it. Partly for this reason, we also provided a list of ineffective practices: teaching approaches that seem to be popularly endorsed by at least some teachers, but whose use is not supported by research (p22).

How teaching leads to learning is undoubtedly very complex. It may be that teaching will always be more of an art than a science, and that attempts to reduce it to a set of component parts will always fail. If that is the case then it is simply a free-for-all: no advice about how to teach can claim a basis in evidence. However, the fact that there are some practices that have been found to be implementable in real classrooms, and that implementing them has led to improvements in learning, gives us something to work with. Much of this work is under-theorised and difficult to make sense of. However, the Dynamic Model of Creemers and Kyriakides (2006) provides a theory that is well specified and has withstood some credible attempts to test it. For now at least, it is the best theory of effective pedagogy we have.

Evidence about methods of evaluating teaching quality

The rise of accountability pressures in many parts of the world have led to a big growth in the desire to evaluate the quality of teaching. A number of methods have been widely used and evaluated in research studies.

Value-added models are highly dependent on the availability of high-quality outcome measures. Their results can be quite sensitive to some essentially arbitrary choices about which variables to include and how to fit the models. Estimates of effectiveness for individual teachers are only moderately stable from year to year and class to class. However, it does seem that at least part of what is captured by value-added estimates does reflect the genuine impact of a teacher on students' learning.

Classroom observation seems to have face validity as an evaluation method, but the evidence shows that the agreement between different observers who see the same lesson is not high; neither is agreement between estimates of teaching quality from lesson observation and from other methods. Levels of reliability that

are acceptable for low-stakes purposes can be achieved by the use of high-quality observation protocols, use of observers who have been specifically trained – with ongoing quality assurance – in using those protocols, and pooling the results of observations by multiple observers of multiple lessons (Strong et al, 2011, Mihaly et al, 2013).

There is some evidence that principals' judgements about teacher quality have positive but modest correlations with other evidence. Inferring the quality of teaching and learning from looking at artefacts such as student work, marking or lesson plans, or from assessing teacher portfolios, is not currently supported by research as valid.

Evidence about developmental use of evaluation

The assessment of teaching quality need not necessarily have summative evaluation as its aim. Indeed, our focus in this review is primarily on formative uses of assessment. In designing systems to support such uses, we need to take account of the characteristics of feedback that are most likely to lead to positive effects and of the environment in which the feedback is given and received.

Specifically, feedback should relate performance to clear, specific and challenging goals for the recipient. It should direct attention to the learning rather than to the person or to comparisons with others. Feedback is most likely to lead to action when it is mediated by a mentor in an environment of trust and support. Sustained professional learning is most likely to result when the focus is kept clearly on improving student outcomes, when there are repeated and sustained opportunities to embed any learning in practice, when the promotion of new thinking about teaching takes account of existing ideas, and when an environment of professional learning and support is promoted by the school's leadership.

A general framework for teaching quality

A number of frameworks for conceptualising the elements of effective teaching have been presented. Broadly speaking they include the following components:

2. (Pedagogical) content knowledge

The evidence to support the inclusion of content knowledge in a model of teaching effectiveness is strong, at least in curriculum areas such as maths, literacy and science. Different forms of content knowledge are required. As well as a strong, connected understanding of the material being taught, teachers must also understand the ways students think about the content, be able to evaluate the thinking behind non-standard methods, and identify typical misconceptions students have.

5. Quality of instruction

Quality of instruction is at the heart of all frameworks of teaching effectiveness. Key elements such as effective questioning and use of assessment are found in all of them. Specific practices like the need to review previous learning, provide models for the kinds of responses students are required to produce, provide

adequate time for practice to embed skills securely and scaffold new learning are also elements of high quality instruction.

4. Classroom climate / relationships / expectations

Again, the empirically based frameworks all include something on classroom climate, though this heading may cover a range of aspects of teaching. Some (e.g. CLASS) emphasise the quality of relationships and interactions between teachers and students. Also under this heading may come teacher expectations: the need to create a classroom environment that is constantly demanding more and never satisfied, but still affirming to students' self-worth and not undermining their feelings of self-efficacy. Promotion of different kinds of motivational goals may also fit here, as may the different attributions teachers make and encourage for success and failure (e.g. fixed versus growth mindset, attributions to effort and strategy rather than ability or luck). Related to this is the valuing and promotion of resilience to failure (grit).

3. Behaviour / control / classroom management

All the empirically based frameworks include some element of classroom management. A teacher's abilities to make efficient use of lesson time, to coordinate classroom resources and space, and to manage students' behaviour with clear rules that are consistently enforced, are all relevant to maximising the learning that can take place. These factors are mostly not directly related to learning; they are necessary hygiene factors to allow learning, rather than direct components of it.

1. Beliefs (theory) about subject, learning & teaching

The idea that it matters why teachers adopt particular practices, the purposes they aim to achieve, their theories about what learning is and how it happens and their conceptual models of the nature and role of teaching in the learning process all seem to be important. Although the evidence to support this claim is not unequivocal, it seems strong enough to include it at this stage.

6. Wider professional elements: collegiality, development, relationships

It seems appropriate to include a final heading that captures some broader aspects of professional behaviour. Danielson's Framework for Teaching includes elements such as reflecting on and developing professional practice, supporting colleagues, and liaising and communicating with stakeholders such as parents. There may not be direct evidence linking these practices to enhanced student outcomes, but if we want to capture a broad definition of effective teaching, they should probably be included.

Best bets to try out and evaluate

Any recommendations we make here are tentative and very likely to be modified. Crucially as well, we must build in robust evaluation into any changes we make; any recommendations are only hypotheses about what might help. Nevertheless, it is important at least to try to capture some suggestions about how we can take these ideas forward to enhance learning. Some actions will be easier than others,

so we have divided them into quick wins and longer term changes. First, though, we outline some general requirements for system improvement.

General requirements

There are a few general requirements that follow from the previous arguments. The first is that a worthwhile system for monitoring and formative evaluation of teaching quality must have at its heart a set of high-quality assessments of student learning. Building in assessment ensures that we keep the focus on student outcomes. If the assessments are of high-quality that ensures that they will capture the learning outcomes that we value and want to incentivise. Ultimately, for a judgement about whether teaching is effective to be seen as trustworthy, it must be checked against the progress being made by learners. However good our proxy measures become, there is no substitute for this.

A second requirement is that a formative teacher evaluation system must incorporate multiple measures, from multiple sources, using multiple methods. Users must triangulate multiple sources of evidence, treating each with appropriate caution, critically testing any inferences against independent verification. The more sources of evidence we have, the better our judgements can be.

A third requirement, related to these two, is the need for a high level of assessment and data skills among school leaders. The ability to identify and source ‘high-quality’ assessments, to integrate multiple sources of information, applying appropriate weight and caution to each, and to interpret the various measures validly, is a non-trivial demand.

A fourth and final requirement is the need to balance challenge and acceptance. If the gap between research-based ‘effective practices’ or data from performance evaluation and existing perceptions is too big the former are likely to be rejected. On the other hand, if the requirements are perceived to be similar to current practice, nothing will change. The latter would be an example of the ‘we think we are doing that’ problem: teachers take on superficial aspects of a new approach, or interpret current practice as aligned with it, and an opportunity for improvement is lost.

Quick wins

A number of specific recommendations should be possible for teachers to implement quickly and without great cost:

1. Spread awareness of research on effective pedagogy.
   The evidence that has been presented in Section 0 about effective teaching approaches may not be universally known by teachers. We should encourage all teachers to engage with these ideas, to challenge their own thinking and that of their colleagues about what is effective, and to understand the kind of evidence that supports the claims.

2. Use the best assessments available.
   Ultimately, the definition of effective teaching is that which results in the

best possible student outcomes. There is currently no guaranteed recipe for achieving this: no specifiable combination of teacher characteristics, skills and behaviours consistently predicts how much students will learn. It follows that the best feedback to guide the pursuit of effectiveness is to focus on student progress, and that requires high-quality assessment of learning.

3. Use lesson observation, student ratings, artefacts and principal judgement cautiously.
   All these methods have potential value, but all have their problems. If they are done well, using the best available protocols, with awareness of how they can be biased or inaccurate, and with due caution about what inferences they can and cannot support, then they should be useful tools.

4. Triangulate.
   A key to suitably cautious and critical use of the different methods is to triangulate them against each other. A single source of evidence may be suggestive, but when it is confirmed by another independent source it starts to become credible. Having more data can sometimes make people feel overwhelmed and indecisive, but for anyone who truly understands the limitations of a single source, being restricted to that would feel hopelessly exposed.

5. Follow the advice from Timperley (2008) about promoting professional learning.
   Sustained professional learning is most likely to result when the focus is kept clearly on improving student outcomes, when there are repeated and sustained opportunities to embed any learning in practice, when the promotion of new thinking about teaching takes account of existing ideas, and when an environment of professional learning and support is promoted by the school's leadership.

Longer term (harder)

In addition to these quick wins, there are other recommendations that may be harder, take longer or cost more to implement. There are broadly two kinds of approaches here: one focuses on developing the measures we need to evaluate effectiveness robustly, the other on developing the support systems that promote the use of feedback for improvement.

Multiple, multi-dimensional measures

If the measures we need do not exist, it may be necessary to create them. If they do exist, but are not yet ideal for our purposes, it may be necessary to develop them further. If they already exist in a suitable format, then we still need to validate them against our criteria for developmental consequences: does using them as part of a formative evaluation process for teachers lead to improved student outcomes?

Create better assessments

In order to judge the effectiveness of their teaching, teachers need to have access to assessments that reflect the learning they are trying to promote, that are calibrated to allow judgements about expected rates of progress, that cover the full range of curriculum areas and levels, and that are cheap and easy to administer on a frequent basis. Although generally of high psychometric quality, available standardised tests do not routinely meet all these requirements.

It may be that system of crowd-sourced assessments, peer-reviewed by teachers, calibrated and quality assured using psychometric models, and using a range of item formats, could meet this need.

Lesson observation tools

A number of protocols exist for lesson observation, and it may be that the best of them provide an optimal way forward. However, it may also be that their requirements for training are prohibitively onerous or expensive, or that alternatives could be developed that better meet the needs of a developmental focus, that are led and owned by the profession, and that make best use of online communities for video sharing, peer ratings and maximising learning for both observed and observer.

One example would be a simple tool for measuring students' time on task in lessons. Brophy and Good (1986, p360) identify the relationship between 'academic engaged time' and student achievement as one of the 'most consistently replicated findings' in the literature. Giving a teacher this relatively objective measure and allowing them to track its trajectory over time and with different classes, perhaps contextualised against the values that other teachers achieve with similar students, could be an effective way to increase the percentage of time spent engaged in lessons and hence to improve learning.

Student ratings

Again, these instruments exist, so this could actually be quite a quick win. Collecting student ratings should be a cheap and easy source of high-quality feedback about teaching behaviours from multiple observers who can draw on experience of multiple lessons. Although there is evidence of using student ratings to enhance learning outcomes in higher education, their use in schools does not appear to have been evaluated yet.

School-based support systems

Creating systems of support within schools that would allow teachers to respond positively to the challenge of improving their effectiveness is an important task. There are many advantages to a school-led system here: it keeps the ownership within the profession and makes the whole process more straightforward to manage. One danger is that without some external expertise the learning may be limited to what is already available in-house (Antoniou & Kyriakides, 2011). It may also be hard to create high challenge in a peer-to-peer system. Part of the reason for generating objective measures of a range of aspects of teaching effectiveness is that they provide an external check against which to compare.

Mentoring

There are many existing models of school-based professional mentoring, so it should not be difficult to select a small number of promising ones for this purpose and evaluate their impact. Key design issues include creating mentoring relationships characterised by trust and feeling supported, while being sufficiently challenging to provoke change. The difficulties of sustaining real change over a long period should also be addressed in the design.

Lesson Study

Another possible route would be to use a Lesson Study approach. Originally from Japan, it was imported in the United States and the United Kingdom and involves groups of teachers collaboratively planning, teaching, observing and analyzing learning and teaching in 'research lessons'. (Dudley, 2014, p. 1)

In the United States, Lesson Study was found to be one of the two interventions, out of the many hundreds systematically reviewed, to have statistically significant positive effects on the pupils' fraction knowledge in grades 2, 3 and 5 (Gersten et al., 2014). Cajkler et al. (2014) argue that Lesson Study provides four benefits: 'Greater teacher collaboration'; 'sharper focus among teachers on students' learning'; 'development of teacher knowledge, practice and professionalism'; and 'improved quality of classroom teaching and pupil learning outcomes.' (ibid., p. 3).

Dudley (2014) suggests that the reasons why Lesson Study works are that it is a gradual process that places specific learners' needs as a focus for development. It involves an element of collaborative enquiry or experiment between teachers who are trying to solve a problem and that takes place 'in the context of a supportive teaching and learning community'. There is also input from an external expertise. In all studies finding positive effects from the implementation of Lesson Study, a considerable role was played by an agent outside the teacher group that could provide feedback and challenge their views.

As with other feedback programmes Lesson Study faces a number of challenges. Saito et al, (2008) report varied opinions among the faculty members with regard to how to observe lessons. Teacher groups 'also differ[ed] in terms of the types of discussions during reflection', with some focusing more on the teaching process and others on student behaviours. Often senior managers or external experts were not involved. Some argue that experiments with Lesson Study may become a practice of 'the blind leading the blind'. This is not a negligible point, and it is one of the main recent critiques to those professional development approaches emphasising practitioners' reflection without providing them with a solid theoretical framework of reference against which to assess them (Antoniou & Kyriakides, 2011).