Research evidence, policy and practice: reflections on the Ofsted research review on science

John Leach

Abstract The first of Ofsted's research reviews focused on science education. It is broad in scope, including content on school systems, the science curriculum, teaching, learning, assessment and materials – with implications for practice. However, the review's scope does not provide much detail of the research on which it is based. I agree with many of the implications for practice, which reflect directions of travel in the curriculum and pedagogy over the last 20 years or so. However, there will be tensions in implementing the findings, given the climate of prioritising ideology over evidence in setting the conditions for teachers' practice in schools.

Compared with most school subjects there is an impressive amount of research evidence and scholarship, from all four Home Nations of the UK and internationally, about science education. This enables decisions on practice to be informed by both theories and evidence rather than ideology ('what has been shown and/or argued and critiqued' rather than 'my opinion'). Millar *et al.* (2006) discuss the complexities around how educational theory, empirical research and practitioner insights are drawn upon when the terms 'evidence based' and 'evidence-informed practice' are used. The first Ofsted research review (Ofsted, 2021) synthesises argument and evidence in the academic literature, including why teach science, what to teach at what age, how to teach, how to assess progress, how to motivate, how to develop materials, how best to deploy the teaching force in school systems, and how to equip the teaching workforce through initial and ongoing development. Of course, questions about the purposes of science education are about values and can't be resolved through empirical evidence, but they have been debated in the academic literature for many years. Anyone making value decisions about science education should show that they have engaged with these debates and justify their decisions in the light of them.

Everyone involved with school science education should therefore welcome the fact that Ofsted is engaging with this body of literature. There has been some scepticism on social media about how serious their engagement is, suggesting that Ofsted is picking selectively from the literature to back up points of existing ideology. Having read the synthesis, I don't share this view, although (inevitably) the brief treatment of big fields of literature means I don't agree with everything and in some cases would synthesise the field in a different way. Nevertheless, I commend the review for engaging seriously with the field.

Seasoned readers of *School Science Review* will remember curriculum and pedagogical initiatives from the 1980s, 1990s and early 2000s, such as the Children's Learning in Science (CLIS) project, Cognitive Acceleration through Science Education (CASE), Science for Public Understanding at AS (age 17), 21st Century Science at GCSE (age 16), and *Beyond 2000: Science Education for the Future* (Millar and Osborne, 1998). In 2006, the UK's Economic and Social Research Council's Teaching and Learning Research Programme commissioned a research synthesis on science education (TLRP, 2006). Most suggestions for practice coming out of that research review will be familiar to readers of *School Science Review*. People will not be surprised that the research review highlights the dual purposes of science education in educating all future citizens as well as preparing for future specialist study, or the importance of engendering a positive self-image as 'someone who can achieve in science' from the earliest years of schooling for future success. It won't be a surprise to read evidence that outcomes are best when ideas are introduced through the curriculum in a manner coherent with the structure of the science, as well as taking account of learners' starting points – tackling existing or learnt misunderstandings through the teaching approach. Readers will be familiar with arguments and evidence about the different purposes of formative and summative assessment, as well as using assessment to support the process of learning amongst students.

Many of the implications for practice identified in the recent Ofsted synthesis echo recommendations from earlier research syntheses and curriculum projects and attempts to put them into practice through curriculum development and ongoing teacher education. I believe that the Ofsted review identifies themes picked up in previous syntheses of research evidence and initiatives arising from them and should not therefore be viewed as being informed by ideology.

The purpose of the synthesis is to identify findings from the research literature as a basis for further development to build practice. In this article I present a response to the synthesis itself, and consider issues about describing educational practices in science education as 'evidence based'.

How does the Ofsted review see the path from research to practice?

Educational research, including science education research, is often criticised for stating the obvious in technical, impenetrable language. I don't believe that the Ofsted research review sees research evidence in that way. However, I do think that, by presenting research evidence about teaching and learning at a high level of abstraction, the Ofsted review risks reinforcing the view that (science) education research adds nothing to practitioner wisdom and common sense. I'll explain why, drawing on two examples of research and development work.

In conceptualising the relationship between 'theory' and 'practice' in science and mathematics education, Ruthven *et al.* (2009) suggested a classification of insights into 'grand theories', 'intermediate theories' and 'design tools'. Piaget's and Vygotsky's insights about knowledge acquisition, and insights from the history, philosophy and sociology of science and neuroscience, are examples of what Ruthven *et al.* describe as grand theories. Although they can give helpful insights about various educational practices (such as teaching, assessing, and designing teaching materials), they require 'translation' to fulfil their aims. Intermediate theories are developed as 'translations' drawing on one or more grand theories to inform educational practices (such as teaching science, assessing science and designing science curricula and teaching materials).

However, grand theories, and even intermediate theories, are not enough to inform practice without further translation. When should the teacher adopt a 'telling' stance in the classroom, when and how should they use question-and-answer, and when is learner–learner discussion the best approach? Fashions in what is viewed as 'best practice' come and go; I have experienced periods in science education where 'best practice' was viewed as learners 'discovering' scientific ideas through their own investigative work, whereas in other periods 'best practice' was viewed as involving teachers declaiming knowledge to learners through direct teaching. Of course, both these approaches are flawed if used alone. Dylan Wiliam recently captured the absurdity of claiming a single pedagogic approach is 'best':

> *Those who argue for the superiority of direct instruction no matter what it is that you want your students to learn remind me of the drunk looking for his keys underneath the streetlamp, not because that is where he dropped the keys, but because that is where the light is.* (Wiliam, 2021: xix)

Teachers have varied views on the use of talk in science classrooms. However, experience suggests that secondary science teachers approach classroom talk in different ways – and that some approaches are far more successful at addressing learning aims than others. This is backed up by evidence (e.g. Chi (2009) and Alexander (2020)).

In order to bring conceptual and empirical evidence to bear on this issue, my colleague Phil Scott and I developed an intermediate theory that we called 'a social constructivist perspective'. It drew upon cognitive and sociocultural learning theory to inform the design and execution of short science teaching sequences to address conceptually demanding content in the secondary school science curriculum (Leach and Scott, 1995: 2002). Together with Eduardo Mortimer, Phil went on to develop the *communicative approach* design tool (Mortimer and Scott, 2003) that classified classroom talk along an authoritative–dialogic dimension (depending on who is steering meaning-making), and an interactive–non-interactive dimension (depending on how many individual voices are contributing to shared meaning-making). Such design tools are used to inform educational practices drawing on intermediate theories. I am using 'educational practices' widely to include approaches to designing curriculum materials, sequencing ideas in the curriculum, teaching, assessment – all the things that create the environment in which learners operate in schools.

At the authoritative end of the authoritative–dialogic dimension the teacher is controlling meaning-making in the classroom. Learner responses are judged as consistent or inconsistent with the conceptual learning aims of the lesson. When the teacher is adopting an authoritative stance to introduce and reinforce a Newtonian account of motion, it is not appropriate to say that objects stop moving because 'the force has run out'. In such cases, when learners use Newtonian ideas incorrectly, it is the teacher's role to supportively correct learners and model the appropriate use of Newtonian ideas. At the dialogic end of the axis, a discussion is happening where both teacher and learners (or groups of learners) make meaning together. Learners have genuine opportunities to change the flow of discussion. An example of where a dialogic approach might be appropriate is enabling learners to practise applying ideas after they have been introduced through authoritative presentation by the teacher. Going back to the motion example, groups of students might practise using previously introduced concepts to generate explanations for real motion events. If the 'force' does not run out, how come a bullet fired from a gun stops moving? By generating such explanations through discussion, learners can practise using, and thereby deepen, their understanding of Newtonian mechanics.

The purpose of dialogue is to enable learners to practise using new ideas and to speak the language of science with their own voices. At the end of dialogue, the teacher will be able to draw upon what they have heard to re-present the scientific canon authoritatively, drawing upon what they have heard learners say, to show the difference between a Newtonian account of motion and some of the ones heard in discussion. They will then establish and reinforce the school science account of the content, making explicit for learners how it differs from their everyday ideas.

The point is that, within a short teaching activity, *both* authoritative *and* dialogic approaches are *designed into the activity* according to the teaching purpose. In this example, the communicative approach adopted in the classroom relates to the *conceptual content* and the *teaching purpose* of the teaching intervention.

Along with another design tool, 'learning demand' (which was used to plan the content, sequencing and presentation of conceptual content), teaching schemes (i.e. worked examples) were developed. The teaching schemes included both activities (with justified learning aims) and guidance on the mode of teacher talk (mapped against its purpose). The design process is described in Chapter 4 of Millar *et al.* (2006). The next step is to evaluate whether the worked examples met their stated aims. 'Research-informed' and/or 'evidence-informed' interventions are not inevitably successful in meeting their aims, or practical for teachers to use in the classroom. We conducted an evaluation of our designed sequences taught to one group against a comparison group who were taught the same content using the school's normal approach to the topic. The groups were matched by both prior attainment in science, and a topic-specific assessment administered before teaching. No special training was provided for the teachers using the designed teaching sequences. We demonstrated statistically and educationally significant improvements in learners' outcomes on conceptual understanding (consistent with the requirements of the English national curriculum in place at the time). The teaching sequences, informed by the use of design tools, as used by the teachers in our sample, 'worked', in that learners achieved statistically and educationally significant learning outcomes compared with learners following the schools' usual approaches (see chapter 5 in Millar *et al.* (2006)).

Perhaps the best-known example of working from grand theory, through intermediate theory and design tools to produce worked examples that can be evaluated, is the CASE project (the CASE team did not use the terminology of grand/intermediate theory, design tools or worked examples; this terminology was created well after the CASE materials existed). Piagetian grand theory was used by the CASE team to develop an intermediate theory to inform teaching for cognitive acceleration. Piagetian grand theory does not mention teaching; it addresses the natural development of intellectual skills through maturation (akin to the physical development of children through adolescence into adulthood). The team then used their intermediate theory to develop the Pillars of CASE (which I think are their design tools). These were then used to write a worked example to tackle the educational job of accelerating learners' cognitive development through the medium of school science education. The worked example was evaluated and demonstrated to have been successful in achieving its aims.

Such approaches to designing and implementing science teaching are open to the criticism that they reduce teachers to the role of technicians who implement a curriculum created by other more powerful actors in the system. As well as setting the curriculum, these other actors determine the values and purposes of science education. I believe that it is inevitable and appropriate that groups other than science teachers have a say in the purposes of science education and the curriculum that attempts to address those purposes. For example, it is hard to justify a curriculum used as preparation for the study of medicine, dentistry and the paramedical professions that took no account of requirements for future study. And moves towards a curriculum that supports science for future citizens will inevitably draw upon insights from a wider base of people than scientists. However, it is critical that science teachers have professional autonomy and agency to contribute to discussions on the purposes of science education and the design and implementation of approaches to realise those purposes through the curriculum. In the context of policy implementation, Pawson and Tilley (1997) argued that it is *people* that implement written policies, thereby *creating* the policy through its enactment. Assuming that policies will be implemented as written is philosophically and operationally naïve.

In the examples above, no claim is made that the worked examples are the *best* or *only* way to address their stated teaching goals. Rather, they are worked examples of successful approaches to addressing those goals that are informed by evidence, highlighting pitfalls, which, if not addressed, will reduce the effectiveness of the teaching.

The Ofsted research review identifies many features of the above examples as recommendations. For example, the recommendations include the need for both conceptual context ('the content') and pedagogic approach to be considered together, and the need for worked examples in practice settings. My comment is that the review does not make enough of the practical design work that is needed to move from theory to practice. The evaluation of CASE and the work on communicative approach demonstrate that insights from research can be used in science education to make improvements that go beyond practitioner wisdom or common sense. However, both examples involved *serious design work* to move from grand and intermediate theory to worked examples that could be evaluated in the classroom. I believe that the Ofsted review opens a real opportunity for systematic curriculum and pedagogical development, involving practitioners, researchers and policymakers working collaboratively, to make a positive impact on the teaching and learning of science. I hope that charities with interest in science education will choose to support such work.

Does the review do justice to the research evidence, or does it pick evidence to support ideology?

Although it is not a quick read, the Ofsted review is not sufficiently long to give more than a cursory account of all the research in its very broad scope. It is therefore inevitable that picky academics (like me) can point a finger at omissions and oversimplifications. However, there are some areas where I think the review's treatment of issues is sufficiently incomplete as to misconstrue the issue, or where I take a different view. For example, in my opinion the whole document doesn't differentiate sufficiently between curriculum (what is taught) and learning (what is learnt). Just because ideas progress in the curriculum, it can't be assumed that learners' understanding progresses in the same manner. The curriculum proposes a progression model (without research evidence in terms of the age-placement and sequencing of content), and there is ample evidence in science education that learners' ideas don't develop according to that model in some fundamental conceptual areas. And there is a failure in the Ofsted review to recognise a strand in the literature that differentiates between the division of science into physics, chemistry, biology (and possibly Earth sciences and environmental science) in school science, and the way that modern professional science is structured. Although I taught chemistry and biology in schools (and chemistry in initial teacher education), my first degree is in pharmacology and included courses in biochemistry, physiology and psychology, as well as chemistry. Even in the early 1980s my university did not offer a degree in biology, instead going straight down to specialist fields.

This has profound implications for the way that 'cross-curricular' themes are presented in the review. Ideas such as energy and models of matter were created before science was institutionally divided into physics, chemistry and biology. In any case, contemporary science isn't structured like this. The review appears to present fundamental ideas as singularly rooted in a single school science discipline, to be applied across the whole of science. However, each field has developed its own conventions for using fundamental ideas about the structure of matter and the role of energy in generating descriptions and explanations, and this has implications for both the curriculum and pedagogy in different content areas. Think, for example, about the energy profile diagrams used in A-level chemistry across the course of a chemical reaction: these are discipline-specific representations drawing on the fundamental concept of energy.

The treatment of 'misconceptions' in the synthesis is quite general, at the level of intermediate theory. There is a lot of work conducted in the 1980s, 1990s and 2000s about learners' pre-instructional reasoning and how insights from this research programme were built into recommendations for practice (such as some materials for the National Strategies in the early part of the millennium). None of this research is referenced in the review. And although it is not my area, I am certain that people who know about cognitive science and neuroscience would feel that the review does not do justice to evidence from those fields about the nature of learning in general. As far as I can tell, the Ofsted review uses neuroscience to justify ideas presented to me on my PGCE course in the mid-1980s, such as 'sequence ideas logically' and 'don't over-face learners with content'. There must, surely, be more to it than that. I worry that such advice is at the level of intermediate theory and therefore says rather little to teachers about how to sequence and chunk material about a particular topic for specific learners in a way that makes it teachable and understandable. Further design work is required to develop those insights for teaching specific content in different subject areas at a sufficiently small granularity: teaching learners to use the perfect tense in French is different from teaching learners to use Newtonian ideas to describe how objects fall to the ground when dropped.

It is interesting to speculate whether teachers need to know and understand grand and intermediate theories, or design tools, to use worked examples in practice. Our work suggests not (Millar *et al.*, 2006: ch. 5). However, that is different from saying that *if* teachers knew more about the theoretical background of worked examples, their professional decision-making would be improved. The evidence presented in the above chapter suggests that this may be the case.

Some of the suggestions for practice in the Ofsted review, that are consistent with the evidence, strongly contradict recent views on 'best practice' and I don't therefore think the review can be accused of cherry-picking the evidence. Examples include:

- **Pedagogy:** The review spotlights evidence supporting the need for mixed teaching approaches according to the learning objective. For presenting new ideas an authoritative approach might be best where the teacher explains new ideas. But all the evidence is that this is not enough: learners also must practise using ideas and scientific language through discussion, group work, practical work and so on. In contrast, the current view of 'best practice' from some sources appears to be that learners should listen to the teacher explaining target knowledge with authority. Group discussion is viewed with suspicion.
- **Practical work:** The review summarises evidence about what practical work is good for, but also highlights the many ways in which it is used that are ineffective (such as teaching 'processes' without a context, or teaching conceptual content through investigation). While few would argue that practical work is good for all purposes, I can imagine dissenting voices from some teachers and CPD providers around the implication that a significant amount of practical work in English secondary science education doesn't hit the mark in terms of meeting its learning objectives. Indeed, there is evidence that much school science practical work misses its learning objectives almost completely (e.g. Abrahams and Millar (2008)).
- **Initial and ongoing teacher development:** Throughout the review there is evidence of the prominent influence of subject content on the effectiveness of teaching and learning. Yet the recent Early Career Framework for beginning teachers (DfE, 2019) adopts an almost entirely generic and technicist approach to teacher development in the early career. 'Best practice' involves 'following these rules'. No attention is given to enabling teachers to engage critically and influence discussions about the purposes of (science) education and their implementation through the curriculum. Furthermore, though not inevitable, policy drivers towards school-led initial teacher education often result in severe reductions in subject pedagogy during training.

From research review to implementation: what are the issues?

The Introduction to the Ofsted review states:

> *We will use this understanding of subject quality to examine how science is taught in England's schools. We will then publish a subject report to share what we have learned.*

This is the stage of *translating* research findings into recommendations for practice and is therefore fundamentally important for what goes on in school science classrooms.

As outlined above, I believe that, if taken seriously, implementing many of the recommendations for practice identified in the review will require a fundamental change to the environments in which science teachers, teacher developers, curriculum developers and assessment agencies conduct their work in science education. This will need appropriate resourcing. Furthermore, implementing all these changes in one go at system level – as evidenced in the literature – is unlikely to be successful because of (among other things) the finite capacity of systems and individuals for change. We should therefore encourage and welcome an ambitious but measured and sequenced programme of change to curricula, initial and ongoing science teacher education and development, and changes to external expectations about pedagogy and assessment.

However, we should recognise the challenges ahead. In order to get things implemented Ofsted/HMI, while quasi-independent, nonetheless have to operate with the support of Government ministers. Reflecting on my time in science education research, I think there was a period in the late 1990s and early 2000s where genuine collaboration between practitioners, those responsible for setting curriculum and external assessment frameworks, HMI and science education researchers led to bold and positive developments in science education. For example, the group that produced *Beyond 2000: Science Education for the Future* (Millar and Osborne, 1998) included inputs from practitioners, HMI, the regulator responsible for curriculum and assessment at that time, as well as researchers. Curriculum developments to support science education for future *citizens* as well as specialists (such as 21st Century Science), and some aspects of the National Strategies, were heavily (and, in my view, positively) influenced by this work. 21st Century Science only came about through extended dialogue involving university academics, major charitable trusts, examining boards, the regulator (QCA), practitioners, and a publisher. However, this environment had effectively gone by the end of the 'New Labour' era, and after the Coalition Government was formed in 2010 the work was abandoned (including taking the National Strategy materials off the internet).

The positive environment was short-lived. A self-improving educational system would recognise that such interventions inevitably would not solve all the issues that they set out to address, and set up a systematic programme of evaluation and development to build on successful innovations with a view to improving them.

In recent times, Government ministers have proclaimed on both pedagogy (e.g. that students should be seated in classrooms with the focus on a teacher imparting knowledge) and curriculum (e.g. a 'knowledge-rich curriculum'), with no reference to available research evidence. Implementing some suggestions in this research review will require practice to go in a different direction from recent political imperatives. For example, giving learners opportunity to practise may best be achieved through working in discussion groups practising building explanations, and it is questionable whether learning 'about science' would pass the 'knowledge-rich' test. These are precisely the kind of 'progressive' innovations railed against by certain Government ministers, and yet there is clear empirical and conceptual evidence from the UK and elsewhere supporting the *measured* use of such approaches (alongside authoritative, teacher-led approaches depending on the teaching goal).

I believe the team that pulled this review together made a valiant attempt and commend Ofsted for taking a serious look at research evidence. However, I cannot help observing that the speed at which the review was produced, and the brevity of space allowed, have resulted in some serious weaknesses. It would have been beneficial to subject the report to normal peer-review processes (by people with recognised academic expertise, whether employed in universities or schools) to improve it through constructive challenge. Nonetheless, I wish the team well with the inevitable challenges that lie ahead as they move towards the messy but critically important task of implementation.

Acknowledgements

I am grateful for valuable discussions in the preparation of this article with my colleagues Alex Ford, Dr Leigh Hoath and Ed Podesta from Leeds Trinity University, Emeritus Professors Robin Millar and Jonathan Osborne of the University of York and Stanford University respectively, and Professor Justin Dillon and colleagues at the University of Exeter.

References

Abrahams, I., and Millar, R. (2008) Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, **30**(14), 1945–1969.

Alexander, R. (2020) *A Dialogic Teaching Companion*. London: Routledge.

Chi, M. (2009) Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, **1**, 73–105.

DfE (2019) *Early Career Framework*. Department for Education. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/978358/Early-Career_Framework_April_2021.pdf.

Leach, J. and Scott, P. (1995) The demands of learning science concepts – issues of theory and practice. *School Science Review*, **76**(277), 47–51.

Leach, J. and Scott, P. (2002) Designing and evaluating science teaching sequences: an approach drawing upon the concept of learning demand and a social constructivist perspective on learning. *Studies in Science Education*, **38**, 115–142.

Millar, R., Leach, J., Osborne, J. and Ratcliffe, M. (2006) *Improving Subject Teaching: Lessons from Research in Science Education*. London: RoutledgeFalmer.

Millar, R. and Osborne, J. ed. (1998) *Beyond 2000: Science Education for the Future.* King's College London School of Education. Available at: www.nuffieldfoundation.org/wp-content/uploads/2015/11/Beyond-2000.pdf.

Mortimer, E. and Scott, P. (2003) *Meaning Making in Secondary Science Classrooms.* Maidenhead: Open University Press.

Ofsted (2021) *Research Review Series: Science*. Available at: www.gov.uk/government/publications/research-review-series-science.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage.

Ruthven, K., Laborde, C., Leach, J. and Tiberghien, A. (2009) Design tools in didactical research: instrumenting the epistemological and cognitive aspects of the design of teaching sequences. *Educational Researcher*, **38**(5), 329–342.

TLRP (2006) *Science Education in Schools: Issues, Evidence and Proposals*. Teaching and Learning Research Programme. Available at: www.stem.org.uk/resources/elibrary/resource/27208/science-education-schools-issues-evidence-and-proposals.

Wiliam, D. (2021) Foreword. In *The Future of Teaching: and the Myths That Hold it Back*, G. Claxton, pp. i–xix. London: Routledge.

John Leach is Dean of the Institute of Childhood and Education at Leeds Trinity University. Email: J.Leach@leedstrinity.ac.uk