Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

From external evaluation, to school self-evaluation, to peer review

David Godfrey, UCL Institute of Education
david.godfrey@ucl.ac.uk

Introduction

Most modern systems of school education around the world now have highly developed evaluation processes. Many countries have introduced national agencies tasked with measuring indicators of educational quality, such as school inspections or test-based accountability. Following quickly in the wake of external evaluation (EE) policies have followed calls for schools to develop their own capacities for self-review.

An OECD report (2013) describes a number of ways in which developing school evaluation capacity should be a priority for school improvement. Among the report's suggestions are:

- Strengthening school principals' capacity to stimulate an effective school self-evaluation culture
- Promoting the engagement of all school staff and students in school self-evaluation and
- ***Promoting peer learning among schools*** [emphasis added]
  (OECD, 2013, pp.469-470)

On this last point, the OECD report points out that peer learning is particularly useful in systems where schools have a high degree of school autonomy, in order to prevent them from forming an introspective and defensive culture. Through partnerships, groups of schools can stimulate collegial networking, peer exchange, sharing and critiquing of practice, and fostering a sense of common direction *(ibid, p.470)*.

Partly, this movement towards peer learning can be seen in the context of school systems that have matured from ones dominated by top-down external inspection towards ones with increased professional or lateral forms of accountability. Another aspect is the growth of networked approaches to school improvement in many countries. Peer reviews can be seen as an essential part of a network's own evaluation and improvement strategy, having the potential to drive individual school improvement and also to support network level outcomes.

In England for instance, we can see how peer review fits very well into the drive towards a so-called self-improving school system (SISS). The SISS signifies a system of sustainable improvement of schools with a strong focus on bottom-up approaches and locally embedded activities. The conditions of such a system are:

- a structure of schools working in clusters and partnerships to promote improvement
- a culture of constructing and implementing local approaches for improvement; addressing topics that are relevant for a specific locality
- highly qualified people who act as system leaders in creating new knowledge, disseminating knowledge and bringing schools together in partnership work (Hargeaves, 2010).

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

Peer review relationships offer solutions within this framework and also balance out the perception of an overbearing top down external inspection or accountability framework. This appeal to the UK Government from the General Secretary of the National Association of Head Teachers in England makes such a point (see also ch.x for more on the Instead peer review model):

Our education system has transformed itself from the era when Ofsted was first designed, and the Committee should look at the need for Ofsted to change too. Inspections should not be adversarial; they should be a constructive dialogue between the inspectorate and professionals. 80 per cent on England's schools are rated good or outstanding, so NAHT believes the time is right to see greater peer-review within the inspection framework. Our own Instead project offers one such model that would help do that; something we hope the Committee will consider in its inquiry (Russell Hobby, General Secretary NAHT Education Select Committee, 02/03/2016).

Peer review outside the school sector

Peer review has been an integral part of many professional fields outside the school sector. A basic definition of peer review is:

"the evaluation of work by one or more people with similar competences as the producers of the work (peers). It functions as a form of self-regulation by qualified members of a profession within the relevant field" (Wikipedia, accessed, 6th June, 2019).

Peer review is used in accounting, law, engineering (e.g., software peer review, technical peer review), aviation, and even forest fire management (ibid). There is a long history of peer review in Higher Education (e.g. Harman, 1998) and in Further Education, peer reviews have been promoted across Europe (e.g. Gutknecht-Gmeiner, 2013) and largely precede their use in the mainstream schools sector.[1]

Examples of peer review activities across a range of sectors includes:

- Opening up (e.g. government) policies to scrutiny of others (other country representatives). These may include visits by teams to the host country to evaluate the success of an initiative

- Local area services (e.g. Martin and Jeffes, 2011) or institutional (e.g. schools/colleges) quality assurance in which teams of colleagues from equivalent services in other areas visit to identifying weaknesses, or validate existing good practice

- Submitting scholarly work (e.g. a journal article) to the (usually double-blind) review of scholars deemed to be academically qualified to make a judgement about the quality and suitability of this work

- The evaluation of professionals' practice by other professionals in the same practice, e.g. clinical peer review of heath care professionals.

---

[1] One scheme in Hampshire colleges in England has been running since 1993: http://www.eqr.org.uk

Professional bodies also recommend protocols and principles for the conduct of peer review, such as those offered by the Dental Defence Union (DDU) who encourage small groups of dental professionals to work together to improve the quality of service, by reviewing aspects of practice, sharing experiences and identifying areas for change.[2]

Looking across the range of uses in different professions, a number of reasons are given for peer review:

- To evaluate the performance of professional practice, policy or initiative by those qualified in the field to do so

- To decide if work meets the necessary standards required (e.g. for publication or funding)

- To ensure that colleagues' work is of a sufficient standard and thereby protect the status of the profession and one's own risks from sanction (Edwards and Benjamin, 2009)

- To share ideas, learning and plans for improvement among groups of professionals in order to improve 'client' outcomes (e.g. patients)

- To prepare for an external evaluation, inspection or audit.

The process of peer review, particularly how it is applied to scholarly uses, has also undergone much scrutiny and criticism. For example, there are concerns that blind or anonymous review does not work because the reviewers can easily guess the identities of (particularly reputable) scholars, so reviews are not really blind. Furthermore, peer review have long been criticised for the tendency toward confirmatory bias, meaning that reviewers judge the standards of scholarly work more according to established knowledge and reject work that falls outside this (e.g. Mahoney, 1977). Peer review of scholarly work can also disfavour the work of minority groups or lead to nepotism (Wenneras and Wold, 2001). The judgement of peer reviewers in science has also been shown to be highly unreliable meaning that the probability of receiving research grants is due to little more than chance (Cole et al, 1981).

Criticisms of cosy relationships, lack of rigour, subjectivity and insufficient evaluation skills of peers can also be applied to school peer review, and therefore require consideration in its implementation.

Outline of the chapter

This chapter begins by examining peer review in in relation to the general research base on internal evaluation (IE). The emergence of IE in response to and in relationship with, external school evaluation and other accountability measures is examined. Internal evaluation is defined, its rationale explained and, drawing upon relevant research the positive impact and the unintended effects of IE are shown. In addition, the conditions for effective IE are described along with what the research has shown about the interaction of IE and EE.

---

[2] https://www.theddu.com/guidance-and-advice/guides/clinical-audit-peer-review-and-cpd

Following this, the chapter looks at the emergence of peer review as a form of internal evaluation. Peer review in the school sector is defined and described, delineating it from other activities. This is followed by an explanation of the different accountability relationships in peer review activities compared to external evaluations and inspection. The chapter then looks at the current research base for peer review in schools, as well as a major European initiative in the Further Education sector. The main effects and impact of peer review are presented as revealed by current research. Finally, attempts to introduce standards and accreditation for peer review in the English school system are described. I outline what the research has so far shown about the effects for peer review and what we know so far about the conditions needed for effective peer review.

From external to internal evaluation

Many countries have developed strong external evaluation regimes to hold their schools to account and to promote improvement. This is despite a dearth of evidence (especially outside of the UK) that inspection has an overall beneficial effect on the school system with some reported benefits balanced by unintended negative consequences, such as gaming (e.g. Nelson and Ehren, 2014). The perceived imposition of external standards on the teaching profession and on school leaders, alongside the need to be prepared for external evaluations, often with high stakes, has led to a strong drive to promote school self-evaluation (SSE). Indeed, as will be discussed below, many proponents of external evaluation have suggested that the introduction of internal evaluation would follow naturally from the introduction of inspections and that these could be complementary and/or an expression of the maturity of the system (e.g. Hargeaves, D 2012, Barber, 2004). Thus, in the late 1990s many international school systems began to focus their attention on SSE, including most of the more developed European countries, Canada, New Zealand and Australia (Hofman et al, 2009).

In the case of England, the creation of the Office for Standards in Education (Ofsted) in 1992 formed the vision of the then Conservative Government for a strong central external system to hold schools to account. This was coupled with a range of other measures that allowed parents to see public data about all schools, allowing the media to convert these into performance tables of examination results. In response to this powerful set of external accountability measures a study was commissioned by the National Union of Teachers (NUT) into self-evaluation and later published in a book called *Schools Must Speak for Themselves* (MacBeath et al, 1995). The incoming Labour Government viewed SSE very positively, and Ofsted endorsed a framework for self-evaluation inspired by the lead taken in Scotland. This was followed by the publication of a range of quality indicators and sources of data that allowed schools to compare themselves to national benchmarks, at school and individual pupil level, including value-added data (MacBeath, 2005). SSE was specifically promoted at system level and with the support of the external inspectorate, with the aim of allowing schools to make targeted school improvement efforts based on comprehensive outcome data. The SSE was designed to mirror that used by the inspectors and thus form a sequential process of school improvement, initiated by the school itself.

Most OECD countries now have varying degrees of legal requirements in place for schools to conduct self-evaluation (Shewbridge, 2013). Some authors, such as MacBeath have long advocated strongly for SSE in all school systems in order to unleash systemic improvement (e.g. MacBeath, 2005). Although most systems that employ self-evaluation routinely have some kind of SSE followed by school inspection, other countries such as the USA with data/test-based accountability models have shown interest in SSE approaches. This is because they focus more on the processes that drive school improvement and a wider range of educational indicators than test-based accountability alone (Ryan et al, 2013).

The notion of self-evaluation relates to a number of related concepts found not only in education but in other sectors. MacBeath (2006 p.4) outlines these, noting that terms convey particular stances of national systems or summative/formative intentions, such as:

Audit: these suggest taking stock of resources and are summative.
Quality assurance: these are a systematic, and usually external-conducted type of audit, carried out for accountability purposes.

Self-review: these can be synonymous with self-evaluation and often used to indicate a summative over-view e.g. of a whole school rather than targeted areas.
Self-assessment: this may involve an examination of the knowledge, skills and attitudes gained by pupils and can be both summative and formative.
Inquiry or Appreciative Inquiry: used more in North America, Inquiry or AI focus on how an organisation can evaluate its strengths within its own frame of reference. This is essentially formative in nature.
Research: this is sometimes used as a synonym for inquiry (or enquiry in the UK context). This can be formative or summative, often involving a range of stakeholders including student and teacher researchers. (See 'research-engaged schools', for example Godfrey and Brown, 2019).
Self-evaluation: this is meant to be a formative process, embedded into cyclical school practices and linked to pupil learning and achievement.

Another way of characterizing the distinction between the summative or formative purposes of these types of evaluations is made in MacBeath's (2005) comparison of self-inspection (summative) with self-evaluation (formative). Self-inspection (a term first found in Ferguson et al (2000)) is top-down, a one-off event providing a snapshot, is accountability focused and based on a rigid framework and pre-determined criteria. Such a process tends to be risk averse and intent on showing how the school meets its standards rather than how it could exceed them or work on its weaknesses. Self-evaluation is, by contrast, bottom-up, continuous, provides an evolving picture, is flexible, creates relevant criteria and focuses on improving teaching and learning (MacBeath, 2005, p.45). This distinction is also helpful and relevant in the consideration of peer reviews, since there is a risk for visiting reviewers of adopting an inspectorial approach with the consequence of limiting professional dialogue and opportunities for school development (see Ch 5).

Internal evaluation in schools

Defining internal evaluation

The broader term internal evaluation (IE), while often used synonymously with SSE, can be used to make a clear distinction between an accountability process driven from the outside (often a government agency/inspectorate) and one driven from the inside (such as a school or school network). While staff external to the school may assist internal evaluations, these are at the behest of the school leadership and focus on areas that are pre-determined by them. Therefore, school peer reviews are best described as a form of internal evaluation, albeit some authors described them as external evaluations due to the visit of staff outside the institution (e.g. Stinton, 2007; Gutknecht-Gmeiner, 2013). The section below further clarifies definitions of school peer review as described in the chapters of this book, including this slightly grey area. However, to begin with, I use internal evaluation as the umbrella term for both single school self-evaluations and peer reviews conducted over multiple institutions or sites. Much of the research focuses specifically on SSE, however, given the overlap in these processes and sometimes synonymous use, lessons can be drawn from existing research in this area that apply as well to peer review. Later I focus more on the incipient research on peer review specifically, as indeed do the succeeding chapters in this book.

The effects of internal evaluation

A systematic review of the literature on internal evaluation by Nelson, Ehren and Godfrey (2015) looked at the effects of internal evaluation in schools. The report, synthesising

research from 1998-2015, considered empirical evidence from 20 countries. Below I summarise the findings from the synthesis in terms of: the beneficial impacts of IE and how to ensure these are maximized; the unintended effects of IE and the conditions for effective IE.

The positive impact of internal evaluation

The international evidence shows that school internal evaluations have led to a number of positive changes such as:

- increased reflection on school quality and intentions to improve
- greater sensitivity to areas in need of improvement, informing goals and actions for improvement planning
- the identification of professional development needs
- greater ownership of change
- increased professional learning
- revised content or organisation of the curriculum
- the provision of targeted support for groups of pupils (Nelson et al, 2015)

Student achievement is maximized when the internal evaluation is more accurate and school improvement priorities are very specific; when student under-achievement is identified and targeted for improvement and where it is used to drive professional development. Changes to teachers' practices are more likely when schools use an enquiry-based protocol to examine student data and use this to foster the acquisition of the teaching skills and knowledge needed to raise student attainment. Finally, positive outcomes from IE occur when teachers attribute student achievement to their own teaching rather than to external causes, increasing collective teacher efficacy (see Ch 12). Internal evaluations are less successful when teachers and leaders are not supported to use or implement IE and use informal methods, when school leaders are unable to interpret data accurately, and when no time is set aside to interpret or act upon the data collected.

Unintended effects of internal evaluation

As well as the many desirable outcomes from IE, there are also some negative consequences that occur when it is not implemented under the best conditions. The 2015 review points to examples when IE can lead to a fixation on measurement and performativity and compliance rather than improvement. An over-reliance on test-based accountability in IE can also lead to the neglect of other achievements and priorities for evaluation. Staff can also suffer from stress, anxiety, and an increased workload.

Conditions for effective internal evaluation

Given the above potential positive effects of IE, and what is known about potential negative outcomes, the following conditions for effective IE emerge from the literature:

- School staff need to develop evaluation literacy, i.e. how to use research-related, enquiry skills
- IE needs to be properly resourced – especially using effective, validated tools for data collection and adequate time for analysis
- Leadership should focus on the development of an enquiry-oriented culture and endorse the importance of IE.

- District level (or middle-tier/network) support and guidance should be given along with expectations for IE and enable sufficient time to conduct it
- External partners can offer critical friendship and support (e.g. universities/school improvement specialist partners)
- A climate of trust needs to exist and/or be developed in the school and between partners.

The last two points taken from the Nelson et al review are particularly relevant to peer reviews, where trust is needed to share data and work collaboratively across multiple schools and where external support is required to guide the process of reviews.

The interaction of internal and external evaluation

The wider context of the accountability system also needs to be taken into account when considering internal evaluation. For instance, some systems (such as England) have very high stakes external evaluations and this may lead to the 'rehearsal' model of IE as schools feel the need to be in a state of readiness for external inspections.

Janssens and Van Amelsvoort (2008) conducted an exploratory study into the effects of School Self-Evaluation (SSE) used by eight Education Inspectorates in seven European countries: England, Scotland, the Netherlands, Northern Ireland, Denmark, Belgium and Germany (Hesse and Lower Saxony). A research team from the Netherlands inspectorate analysed documents provided to them by each inspectorate. The degree to which schools were given guidance on how to complete the SSE was looked at, as well as the position of the SSE in school inspections. The study explores the extent to which each system orients towards an accountability orientation (AO) or the improvement-orientation (IO) in their SSE. In countries with an improvement orientation (supportive), the SSE occupies a weak to moderate position in the inspection process. Where the national system of inspection has an accountability orientaton (AO), the SSE occupies a stronger position. A key aspect is the amount and type of steering given to schools about the SSE process and framework. The authors say that there is a growing tendency in Europe to use SSE for accountability purposes and that this may conflict with the SI aims.

Some evidence shows how external school evaluations can also strengthen internal evaluations and therefore increase their potential for school improvement. Survey research in six European countries looked at direct and indirect links between the external evaluation, SSE and school improvement. Ehren and colleagues' research showed that the role of external evaluations was to set standards for what constituted a 'good school' and to sensitize stakeholders (parents, principals, students) to external evaluation reports. Both of these in turn, led to improvements in the school's SE. The external school evaluations can stimulate SSE and this can be related to specific school improvement actions (Ehren et al, 2013). These actions concern building capacity by improving teacher participation in decision-making, improving teacher co-operation and also by improving transformational leadership. These in turn relate to improvements in school effectiveness as measured by improved opportunities to learn and assessment of students and the school. These improvements in SSE, more than the acceptance of the principle of the external feedback, were found to be the key to driving improvements.

There are three overall ways to connect internal evaluation with external evaluation (Kyriakides and Campbell, 2004). These are:

1. Parallel existence: The external evaluation is more concerned with accountability and internal evaluation with school improvement
2. Sequential: The school conducts its own evaluation and then an external body uses it as a basis to conduct its own or vice-versa
3. Cooperative models: Internal and external evaluators discuss and negotiate criteria. Measurement criteria are combined in a holistic evaluation, taking into account the interests of all parties.

For the first two models - parallel and sequential -, the external evaluator should be responsible for the accountability agenda. These would tend to centralise the process, towards agencies acting for the government. In the parallel model (such as in the USA), schools would tend to conduct more improvement oriented SE, while the external evaluation is more test-based and accountability driven. The two are quite distinct and can be compared. In the sequential model (such as in Hong Kong), the external evaluators can validate or challenge a school's IE. However, the process can also work the other way, with recommendations by external evaluators being fed forward into the school's IE.

Many school systems may encourage or mandate the use the external evaluation criteria in their own IE. While this may align the two concerns somewhat, the external criteria may lack the context specificity required to drive genuine school improvement. Few systems have the cooperative model which, while possibly an ideal one, may require a degree of resourcing on the part of external evaluators that is hard to achieve in practice. However, the area-based inspections in Northern Ireland could be considered one such example[3]. Kyriakides and Campbell (2004) argue for a maturity model, where the contribution of the school to the process may depend on how well they have performed, moving from the first two models - parallel to sequential - to the third model of cooperation.

An OECD report (2013) suggests that school systems should set up external evaluation in ways in which internal evaluation is optimised and vice-versa. The report cites weaknesses in relying on either alone, since external evaluation can lead to game playing and internal evaluation can be subject to 'self-delusion'. They suggest a number of potential ways in which external and internal evaluations can be mutually reinforcing. While schools best know their own contexts, external evaluators can provide rigour and expertise in interpretation and validation of the school's judgements. The OECD report describes a number of ways to increase this coherence at system level, including: developing agreed national criteria on school quality; developing appropriate resources for schools to use in their self-evaluations; ensuring a strong evidence base for external school evaluation and appropriate analysis tools; and ensuring transparency in external school evaluation procedures.

From within school self-evaluation to peer review

So far this chapter has looked at internal evaluation more broadly and its relationship with external evaluation in school improvement. Much of the research looks at within school self-evaluation, although SSE is also an integral part of a sequence of IE that can lead to peer review.

---

[3] http://www.schoolinspections.eu/update-after-one-year-studying-school-inspections-in-northern-ireland/

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

In the schools sector, the overlap with peer observations, networked learning communities, learning rounds and collaborative enquiry means that more attention is needed to adequately describe and delineate peer review activity. Thus, I outline such a definition, below:

School peer reviews are evaluations carried out by peers of schools or parts of schools (such as departments, subject areas or year groups). Schools nominate staff to collaborate with other schools in networks, partnerships and clusters to collect and analyse data in school review visits. These visits usually build on the school's own self-evaluation in the area of focus for the review, offering validation or challenge to the school's own findings. Visiting review teams provide feedback to the school (initially verbally) and often a report is produced for the school's internal use to summarise the findings and to give recommendations.

In most peer review programmes, there is usually some form of mutuality or reciprocity involved, in that schools may sign up to be visited but also reviewers from this school may visit another one. In some larger organisations that conduct peer review (e.g. Challenge Partners) this mutuality may be spread out among members of a larger network, so that schools can call upon peers to conduct reviews of specific areas or their own staff may be called upon to visit another on request. The mutual learning also comes from school visits, where reviewers are learning new practices and policies while they are evaluating the host schools.

School peer reviews provide feedback, critical friendship and validation (or not) of the school's self-evaluation. Peers can also support fellow schools' improvement efforts. While peer review programmes are often structured, facilitated and accredited by external agencies (e.g. EDT, Challenge Partners, UCL Institute of Education), many schools, local authorities and networks set up their own programmes and devise their own cycles and schemes for peer review. The peer review activity is likely to be supported and structured within a network (e.g. Challenge Partners, see chX), a school improvement programme (e.g. the Schools Partnership Programme, see chx) or it can sit within an evaluation cycle of a broader and permanent alliance of schools[4].

As with other forms of internal evaluation, peer reviews can be more improvement oriented (formative evaluation to 'improve') or more accountability oriented (summative evaluation to 'prove'), although the (at least stated) intention of peer reviews is unarguably more the former. For instance, while external evaluation criteria may be used in the conduct of school peer reviews, there is no obligation to publish reports from peer reviews as their purpose is to help the school in its improvement efforts.

Peer reviews are not the same as:

- Internal school peer observations or other within-school peer coaching or learning

---

[4] Listen to this example of a Teaching School Alliance in England on how they incorporate peer review: https://audioboom.com/posts/6854494-matt-davis-and-marie-claire-bretherton-talk-school-to-school-collaboration

- Networked learning communities or research learning communities (Brown 2017), which may involve data analysis and collaborative school enquiry but do not involve evaluative school site visits

- Inspections - even in systems where these are entirely or mainly carried out by those currently serving as school leaders. This is because the accountability dimension is different, e.g. holding the school to account on behalf of the Ministry or central government.

Although usually voluntary (at least for school leaders), some systems, such as England, are beginning to embed peer review to the extent that they have become an expected part of the internal evaluation process for many schools. Some tighter networks (such as Multi-Academy Trusts (MATs) in England) may mandate peer reviews for all the schools in their trust. In these cases, the peer reviews are still 'internal evaluations' as they are internal to the trust, albeit may be seen as external to individual schools.

The emergence and growth of peer review in school systems.

Recent incarnations of peer review programmes have emerged in response to perceptions among school leaders and others, about the dominance of external accountability. The notion of peer review suggests a relationship in which the knowledge and perspective of each party is equally recognized. Davis and White (2001) suggest that peer review in a sense was more prevalent up to the end of the 1970s, when a professional model prevailed that:

.. assumed that educational quality was best ensured by trusting teachers, advisers and others with relevant training to make decisions in the interest of pupils. In this model professionals decided on 'good practice' (Davis and White, 2001, p. 675).

The authors give examples of how local education authority advisors and HMIs assisted in school evaluations, going on to say that the school system (in England and Wales) has subsequently become much more centralized and standardized, reducing the power of teachers and school leaders to make their own decisions. For Barber (2004), there was a shift from informed prescription in the 1990s to informed professional judgement in the 2000s. He suggested that the locus of responsibility would shift more onto teachers and school leaders in a more mature system. While not specifically promoting peer review, he said that this period would need to include sharper, more intelligent forms of accountability.

An OECD review in 2013 reported that peer review practices were emerging across several countries, including the Czech Republic, Finland, England, Sweden and provides a case study of multiple peer review practices in Belgium. The latter case, published in an earlier OECD review (Shewbridge, 2011), gives examples of how schools have reported increased skills in critical friendship and in self-evaluation capacity through their involvement in collaborative peer review networks.

Although reliable data elsewhere is not available, peer review in the schools sector is probably more established in England than elsewhere in the world. In a 2019 think piece, Gilbert notes that peer review is increasingly part of local area partnerships' change strategies and school improvement work in England (Gilbert, 2019). As pointed out by Greany in Chapter 4, a survey conducted in 2018 showed that nearly half of all schools had engaged in peer review in the previous year (Greany and Higham, 2018). Challenge Partners was a pioneer organisation, emerging from the London Challenge programme. Running

from 2002-11, this programme emphasised strong system leadership and partnership working with the creation of new roles and of designated Teaching Schools (see Ch 10).

Other big networks have subsequently formed in England, notably the Schools Partnership Programme (SPP, see Ch 11). Recent survey data, obtained from an ongoing evaluation by the Education Endowment Foundation (EEF) of the SPP shows this trend towards growth in uptake of peer review in schools in England. Of the 339 primary schools surveyed in June 2018, a third of the sample (111 schools or 33%) said that they had been involved in a peer review programme other than SPP over the two years prior to the survey. Of this third, 59% said that this was a model developed by themselves in partnership with other schools. There were also a variety of other sources for peer review, including 14% were using a local authority model or one used in part of their formal school (Multi-Academy) Trust. Some also mentioned specific providers who specialised in school improvement networks revolving around peer review, for instance Challenge Partners (10%). We should bear in mind all of these schools are taking part in the School Partnership Programme (SPP) and had not done so previously as a condition of the trial (Anders et al, forthcoming). The figures from this sample of English schools show a sizeable minority of schools used to using peer review and this figure seems likely to continue to grow.

Peer review and the accountability dimension

As a form of IE, peer review fits within a wider school accountability framework and may be adding additional layers of accountability to the system above and beyond what other forms of evaluation achieve. Earley and Weindling (2004) outline four accountability relationships in the school system:

- Moral accountability (to students, parents, the community)
- Professional accountability (to colleagues and others within the same profession)
- Contractual accountability (to employers or the government)
- Market accountability (to clients, to enable them to exercise choice)

In many countries, the last two forms of accountability are heavily emphasised. These forms of holding schools to account tend to be taken up by central, external evaluation agencies (such as Ofsted in England) and have a largely summative function. The first two, argues Gilbert (2012) are less emphasised and should have more attention placed on them. These are the forms of accountability that underpin most peer review activity and can be seen for example, in the pledge that schools take when working together in the Challenge Partners' programme (Matthews and Headon, 2015). While the moral dimension provides the overall rationale for working together, this is achieved through the challenge and support provided by peers. This form of professional accountability can be seen as an extension of the kinds of professional responsibility seen as inherent in many professions, notable in medicine, with the so-called hippocractic oath[5]. Gilbert (2012) cites a suggestion by a college principal for teachers' standards, adopted from the medical model which would include: "protect and promote the education of students both within your school and across the schools system", and "working with other colleagues and schools/colleges in ways that best serve the interests of all students" (Gilbert, 2012, p.11). The elements of both professional and moral accountability are clearly exemplified here and peer reviews focus around both aspects when they involve teachers (and not just school leaders) in the process.

Andreas Schleicher of the OECD (2018) emphasises the importance of this form of professional accountability, in which "teachers are accountable not so much to administrative authorities but primarily to their fellow teachers and school principals (p.116). This has also been characterised as a shift towards greater horizontal accountability, where school-to-school networks complement the vertical accountability system of assessment and Ofsted inspection (Grayson, 2019, p.25). In many countries where school networks have been emphasised, peer reviews offer a structure for increasing levels of this kind of accountability and complementing the existing vertical mechanisms.

The research base on school peer review

So far there has been little published research on peer review in schools, so questions of effectiveness, particularly on mechanisms by which they might raise student outcomes, are still unanswered. The rest of this book brings together a collection of work that will add to

---

[5] now updated by the World Health Association and called a pledge, see:
https://www.bioedge.org/bioethics/new-hippocratic-oath-for-doctors-approved/12496

this knowledge base, including several empirical studies. Below I outline some other examples.

One early, published research in the Netherlands studied 27 primary schools between August 2004–June 2006. This looked at how the so-called SVI model was implemented (Blok, Sleegers and Karsten, 2008). This model was specifically designed to balance the demands of external and internal evaluation and avoid the loss of professional voice found in a hierarchical system of evaluation with external evaluators on top and practitioners below. In this model, the schools, divided up into three regional clusters, carried out their own self-evaluation within an agreed framework. This was followed by a one-day visitation from a team of teachers and school leaders from other schools. Schools were encouraged to take part in review visits as well as to host them, in order to create a sense of mutual dialogue and learning. The final phase involved inspectors from the Dutch national inspectorate making a preparatory visit followed by a regular inspection. The former visit was designed to clarify the school's self-evaluation and the visitation report and to set the scope of the formal inspection visit.

School principals taking part in this Dutch research reported significant benefits from the visitation of critical friends, albeit finding it time consuming. They found that the process helped them to build the school's capacity to improve (Blok et al, 2008, p.391). However, Blok and colleagues examined the quality of reports from the SSE, the visitation and the external evaluation. They concluded that the visitation reports were often low in quality or failed to provide empirically based conclusions about the school's own self-evaluation. Interestingly though, in the 24 visitation reports they analysed, eight reports considered the school SE alone, 12 considered the SE plus other questions and four considered only 'other questions'. This was despite guidance given to all schools that ONLY the school's SE should be considered for the visits. The authors concluded nevertheless, that the focus of the visits were almost certainly agreed by the visited school and the visitation team. Thus, it may be that the benefits of the visits were not explicit in the reports themselves and these may reflect an unimportant part of the process to the schools involved. Nevertheless, this research does lend support to the idea that peer evaluations of this nature are not straightforward to either conduct or report. The positive evaluations of the clusters by participating staff were also highly likely to be partly as a result of very committed schools who had self-selected their involvement in the project.

Earlier I mentioned an ongoing trial by the EEF of the Schools Partnership Programme (SPP)[6], this is the largest such trial run so far, involving well over 300 schools in a matched propensity design, comparing the improvement trajectories of similar schools over two years. The study will look at improvements in numeracy and literacy scores compared to the matched schools. It must be emphasised that the SPP is not only a peer review programme but an overall 'package' of school improvement driven by partnerships of schools who self-evaluate, peer review and provide ongoing support and challenge. Nevertheless, the peer reviews form a central plank of the SPP and can be seen to be the heart of what SPP aims to do in terms of increasing professional accountability and providing agency for change to the school leaders and teachers within the partnerships. A large part of this study also looks at the implementation process, in order to tease out the intermediate variables and conditions underlying this school improvement process, following SPP's own theory of action (see also Ch 11).

---

[6] see: https://educationendowmentfoundation.org.uk/projects-and evaluation/projects/schools-partnership-programme-spp/

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

In an independent evaluation of the Challenge Partners (CP) programme (Matthews and Headon, 2015, see also Ch. 10) found a number of positive impacts from peer review (quality assurance (QA) visits). The authors looked at the review system, protocols, handbooks and guidance; they conducted documentary analysis of 25 QA reports and surveyed and interviewed numerous stakeholders. The surveys included 70 headteachers, 71 lead reviewers, 200 reviewers and 43 trainee reviewers. The report concluded that there were 'multiple gains' from the CP programme, not least of these was the high quality professional/leadership development gained by participants. This came about through leading and participating in reviews and also through the training and interchange of professional dialogue. Eighty-four per cent of headteachers of reviewed schools felt that reviews had been very useful to the professional development of their senior leaders and over 90% of headteachers felt that they had helped in planning school improvement. The report set out a number of conditions for successful peer reviews (see below) and also recommendations for how such a programme could contribute at system level.

European Peer Review Network

While not in the (primary/secondary) school sector, extensive work and research has been carried out in Further Education (FE), a sector similar enough form which to draw some useful lessons.

In Europe three phases work on peer review have been implemented and studied in FE and other sectors between 2004-2009. So far, in the three projects (the last one finishing in 2018), 38 project partners from 15 European countries have taken part (Austria, the Czech Republic, Denmark, Finland, Germany, Hungary, Italy, the Netherlands, Portugal, Romania, Switzerland, Slovenia, Spain, Turkey, and the United Kingdom). In these projects, transnational peer reviews involved assembling a team of peer reviewers from at least three different countries to come and visit and conduct the peer review at a host institution sighed up to the project (e.g. Gutknecht-Gmeiner, 2013). These projects focused first on initial vocational education and training (VET) and then adult education and non-formal and informal learning. There have also been a number of transfer of innovation projects, including in Lithuania, Finland, Slovenia and France.

The initial projects were inspired by HE use of peer reviews to evaluate institutions or departments and adapted a model to use in initial VET. The European peer review network was set up to build on these projects and to create an ongoing Europe-wide network of events and partner institutions by the Austrian government[7]. One of the outcomes was to produce tools to measure the impact of peer review in VET and also to devise a peer review assessment tool.

Looking at the impact of the 2004-2009 phases, the research found self-reported improvements in the quality areas implemented in 13 of the 14 case studies, since the peer review took place (Gutknecht-Gmeiner, 2010). The report states that:

Impacts of these improvements as measured for instance by satisfaction rates of different stakeholders (students, staff, cooperation partners like enterprises), better achievement

---

[7] http://www.peer-review-network.eu/pages/about-us.php

rates of students etc., can be detected in half the cases based on quantitative survey results and indicators or similar evidence (Gutknecht-Gmeiner, 2010, p.33).

In several of the case studies, improvements were also reported in other quality areas, suggesting the knock on effect of peer reviews beyond their initial focus. Other changes included changes to quality assurance procedures and evaluation tools as a result of their involvement (ibid, 33-34). Finally, the report notes that, "in several cases, the pilot Peer Reviews have also influenced quality management systems on the national/regional level", including in Scotland, Romania, Austria, Denmark and Finland (ibid, p.35). The last result may reflect some of the natural growth trajectory as mentioned above, mirroring the school sector but also most likely the positive experiences of those that took part in the scheme.

The international cooperation in this project provides a novel dimension that could be usefully applied to the schools sectors as well as the work done to decide on shared tools and frameworks for peer review.

Standards for peer review

In the above European projects, much emphasis was given to the drawing up of processes, handbooks, guidelines, tools and standards for peer review. Indeed these have led to Europe wide standards and have been adopted across many countries, as mentioned above. While the adoption of peer review has lagged behind in the school sector, various organisations have sought to do the same in the English school system. So far, each organisation involved in devising and supporting peer review schemes has drawn up its own standards, however in recent years the NAHT has set up its accountability commission which includes peer review in an overall assessment of the current system of accountability in England.

The commission's concern is that the plethora of models for peer review may hide many examples that lack the sufficient rigour to be effective. Recommendation 7 of the commission suggests the need to evaluate existing peer review programmes to identify characteristics of effective practice in order to develop national accreditation arrangements (NAHT 2018, p.6). The commission observes:

The English education system is on a journey; too few schools currently engage in peer review, and not enough is yet known about the essential characteristics of effective review and the conditions in which it has an impact (ibid, p.19).

The aim of the commission is to further understand the features of effective peer review and then to establish, "accreditation arrangements to oversee the expansion of suitable models that are proven to deliver tangible and sustainable benefits" (ibid, p.19).

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

Conditions for effective peer review

The current evidence based already suggests several conditions under which peer review may be more or less effective, some of these will be explored further in Chapter 13, which looks at how peer reviews could be scaled up into the school national evaluation system. Here I summarise some of the key learning points.

Matthews and Headon's 2015 report highlighted several conditions for effective peer reviews:

- Having well trained, skillful lead reviewers, experienced in inspection methodology[8] and a high quality reviewing team
- Reviews that adapt to schools' needs and desired areas of focus
- A school leadership and staff predisposed to receiving feedback positively
- A willingness to meet the cost and allocate time to reviewers (although the cost compared to other training programmes is likely to be minimal)
- Ensuring that peer reviews are conducted in a developmental rather than inspectorial approach, being open and honest about weaknesses as well as strengths
- A high level of trust across partner schools and willingness to maintain and build on this partnership work
- Sustainability and moving onto multiple cycles of review may be helped by adapting approaches, including focusing on specific areas of improvement within a school or by conducting whole school reviews when schools are 'in-between' inspections, to act as a 'temperature gauge' of current performance.

One of the contributors to the NAHT commission, Kate Chhatwal of Challenge Partners, comments that there are four essential ingredients of peer review programmes[9]:

Independence: the reviews should be led by people who have sufficient distance to give an honest and impartial evaluation of the school or the network. These reviewers should evaluate not according to comparisons with own school but by asking what they are doing and what impact it is having. Reviewers can also gain from the experience and take back ideas to their own setting.

Reviewing 'with' the school: the review is conducted alongside the school's leaders and generates professional dialogue rather than giving a sense that they are being 'done to'.

Revealing the skeletons: schools need to show rather than hide any difficulties they are having from the review team.

Commitment to better outcomes: these should be shared outcomes, for all schools being reviewed not just to their own setting. Schools need to commit to working before and after their annual reviews.

---

[8] this may be balanced by the danger of approaching reviews too much as if they are real inspections (see chX). However Matthews and Headon see as advantageous that reviewers can give an assessment of where the school would stand if inspected (ibid, p.45).

[9] https://www.challengepartners.org/news/comment-power-peer-review

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

Although applied to the CP programme, most of the above comments may be applied to other peer review programmes. To these can be added, from the experience in the European network and in the Dutch research by Blok et al:

- the requirement for ongoing training of review teams to ensure high quality reviews
- and the use of a well thought out and shared framework for evaluation
- the use of high quality (trialed) tools for evaluation

Further research on conditions for effective peer review, as well as those that compared the efficacy of various models in use worldwide, to which this volume contributes, should add to the much needed knowledge base in this growing movement.

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

References

Anders, J., Godfrey, D., Stoll, L., Greany, T., McGinity, R. & Munoz-Chereau, B. (ongoing) Evaluation of the Schools Partnership Programme. *An evaluation for the Education Endowment Foundation*

Barber, M. (2004) The virtue of accountability: system redesign, inspection and incentives in the era of informed professionalism. *Journal of education*, 185: 1-38.

Blok, H., P. Sleegers & S. Karsten (2008), Looking for a balance between internal and external evaluation of school quality: Evaluation of the SVI model, *Journal of Education Policy*, Vol. 23(4), pp. 379-395.

Brown, C. (2017). Research learning communities: How the RLC approach enables teachers to use research to improve their practice and the benefits for students that occur as a result. *Research for All*, *1*(2), 387-405.

Chapman, C., & Sammons, P. (2013). School Self-Evaluation for School Improvement: What Works and Why?. *CfBT Education Trust*. 60 Queens Road, Reading, RG1 4BS, England.

Cole, S., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, *214*(4523), 881-886.

Earley, P and Weindling, D (2004) *Understanding School Leadership*, London: Sage

Edwards, M.T. & Benjamin, E.M. (2009). The process of peer review in US hospitals. *Journal of Clinical Outcomes Management*. **16** (10): 461–467.

Ferguson, N., Earley, P., Fidler, B., & Ouston, J. (2000). *Improving schools and inspection: The self-inspecting school*. Sage.

Gilbert, C. (2017). Optimism of the will: the development of local area-based education partnerships. A think-piece. LCLL publications. Accessed (11th June, 2019) at: http://www.lcll.org.uk/publications.html

Gilbert, C, (2012). Towards a self-improving system: the role of school accountability. National College for School Leadership.

Godfrey, D., & Brown, C. (Eds.). (2019). *An Ecosystem for Research-Engaged Schools: Reforming Education Through Research*. Routledge.

Grayson, H (2019) School accountability in England: A critique. Headteacher Update. Vol 2019 (1), pp.24-25. Published Online: 10 Jan 2019. https://doi.org/10.12968/htup.2019.1.24.

Greany, T., and Higham, R., (2018) *Hierarchy, Markets and Networks: analysing the 'self-improving school-led system' agenda in England and the implications for schools.* London: IOE Press.

Gutknecht-Gmeiner, M. (2013). Peer Review as external evaluation in vocational training and adult education: definition, experiences and recommendations for use. *Edukacja ustawiczna dorosłych*, (1 (80)), 84-91.

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

Gutknecht-Gmeiner, M. (2010). Peer Review Impact Analysis Report. Analysis of 14 transnational European Peer Reviews carried out in eight European countries 2006-2009. Accessed online (12th June 2019) at: https://www.oph.fi/download/130480_Peer_Review_Impact_Analysis_Report_10_11_29_final.pdf

Gutknecht-Gmeiner M. (ed.) (2009), European Peer Review Reader, Vienna. Accessed online (12th June, 2019) at: https://www.oph.fi/download/120561_European_Peer_Review_Reader.pdf

Hargreaves, D. H. (2012). A self-improving school system: towards maturity. Accessed online (13th June, 2019) at: https://dera.ioe.ac.uk/15804/1/a-self-improving-school-system-towards-maturity.pdf

Hargreaves, D. H. (2010). Creating a self-improving school system. Accessed online (13th June, 2019) at: https://dera.ioe.ac.uk/2093/1/download%3Fid%3D133672%26filename%3Dcreating-a-self-improving-school-system.pdf

Harman, G. (1998). The management of quality assurance: a review of international practice. *Higher Education Quarterly*, *52*(4), 345-364.

Hofman, R.H., Dijkstra, N.J. & Hofman, W.H.A. (2009) School self-evaluation and student achievement. *School Effectiveness and School Improvement*, 20 (1), 47-68.

Janssens, F. J., & van Amelsvoort, G. H. (2008). School self-evaluations and school inspections in Europe: An exploratory study. *Studies in educational evaluation*, *34*(1), 15-23.

Kyriakides, L. & Campbell, R. J. 2004. School self-evaluation and school improvement: A critique of values and procedures. *Studies in Educational Evaluation,* 30, 23-36.

MacBeath, J. (2005). Self evaluation: Background, principles and key learning. *Nottingham: NCSL*. Accessed online (3rd June 2019) at: https://dera.ioe.ac.uk/5951/3/self-evaluation-background-principles-and-key-learning.pdf

MacBeath, J. (2006). *School inspection and self-evaluation: working with the new relationship.* London: Routledge.

MacBeath, J., Boyd, B., Rand, J., & Bell, S. (1995). *Schools speak for themselves.* Accessed online (13th June, 2019) at: https://www.teachers.org.uk/files/active/0/schools.pdf

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive therapy and research*, *1*(2), 161-175.

Martin, K. and Jeffes, J. (2011). Safeguarding Children Peer Review Programme: Learning and Recommendations. Slough: NFER.

Matthews, A. P., & Headon, M. (2015). *Multiple gains: An independent evaluation of Challenge Partners' peer reviews of schools*. Institute of Education Press.

Please note. This is an earlier draft to published version and may contain several errors and page numbers do not correspond

Michek, S., Pol, M., & Chvál, M. (2013). Peer review in school self-evaluation: Cultivating skills in Data use. *The New Educational Review*, *34*(4), 30-40.

NAHT (2018) Improving Accountability. *Report of the NAHT accountability commission. National Association of Head Teachers*. Accessed online (7th June 2019) at: https://www.naht.org.uk/news-and-opinion/news/structures-inspection-and-accountability-news/improving-school-accountability/

Nelson, R., Ehren, M., & Godfrey, D. (2015). Literature review on internal evaluation. *London: Institute of Education*.

Nelson, R., & Ehren, M. (2014). Review and synthesis of evidence on the (mechanisms of) impact of school inspections. *University of Twente*.

Nisbet J. (1990), Rapporteur's Report*, In: *The evaluation of educational programmes: methods, uses and benefits: report of the educational research workshop held in North Berwick (Scotland)*, 22-25 November 1988, Scottish Council for Research in Education (Hg.), *[= European Meetings on Educational Research, Part 1, vol. 24*], Amsterdam: Swets & Zeitlinger. (as cited in Gutknecht-Gmeiner M. (ed.), 2009, above)

OECD (2013), *School evaluation: From compliancy to quality*, in Synergies for Better Learning: An International Perspective on Evaluation and Assessment, OECD Publishing, Paris, pp. 383-485. DOI: https://doi.org/10.1787/9789264190658-10-en

QAA (2015) *Higher Education Review: A handbook for universities and colleges with access to funding from HEFCE or DEL undergoing review in 2015–16*. Online. www. qaa.ac.uk/en/Publications/Documents/HER-handbook-15.pdf

Ryan, K.E., Gandha, T., & Ahn, J. (2013). School Self-evaluation and Inspection for Improving U.S. Schools? Boulder, CO: National Education Policy Center. Accessed online [2nd June 2019] at:
http://nepc.colorado.edu/publication/school-self-evaluation

Sahlberg, P. (2011). *Finnish lessons*. Teachers College Press.

Schleicher, A. (2018). *World Class: How to Build a 21st-Century School System*. Paris: OECD Accessed online (11th June, 2019)at: http://www.oecd.org/education/world-class-9789264300002-en.htm

Shewbridge, C., Hulshof, M., Nusche D. & Stoll, L. (2011), *OECD Reviews of Evaluation and Assessment in Education: School Evaluation in the Flemish Community of Belgium*, OECD Publishing, Paris, www.oecd.org/edu/evaluationpolicy

Stinton, J. (2007) *Evaluation of the College/School peer review pilot project*. Report No. RS7007, The Research Centre, CCN, Norwich

Wenneras, C., & Wold, A. (2001). Nepotism and sexism in peer-review. *Women, science and technology: A reader in feminist science studies*, 46-52.