Literature Review on Internal Evaluation

R. Nelson, M. Ehren, D. Godfrey

September 2015

For more information:

dr. Melanie C.M. Ehren
Institute of Education
University College London
20 Bedford Way
London WC1H 0AL
6th Floor, room 665

E-mail: m.ehren@ioe.ac.uk
Phone number: +44 (0)20 7612 6830
Skype: melanieehren

Table of Contents

1. Introduction

This review is part of the EU-project 'Polycentric Inspections of Networks of Schools' (PINS). The purpose of this review was to identify and summarise findings from international empirical and non-empirical research on the impact of internal evaluation in schools. It sought to answer the following questions:

1. What are the *effects* of internal evaluation? For students, does it lead to improved student attainment and/or improved conditions of learning? For the school, does it lead to reflection and intentions to improve and is there evidence of actual school improvement as a result?
2. What are the *conditions* for successful internal evaluation?
3. How do internal evaluations lead to improved student attainment and/or improved conditions of learning? What are the *mechanisms* that underlie effective internal evaluation?

1.2 What is internal evaluation?

The definition of internal evaluation used to guide the review is of *a process of purposive evaluation of school practices which provides insights into the educational experiences of students, as more than those measured by test data* (Simons, 2013). It is variously termed: internal evaluation, (school) self-evaluation, self-review, data use, data- based decision making, inquiry, internal accountability. Internal evaluation is carried out solely by staff internal to the school and these may be groups of teachers and/or other members of staff, school leaders or designated project staff (Nevo, 2001). 'Data coaches' may work, as employees of the school, with teams to facilitate the interpretation of data and in using it to plan changes to teaching approaches. Schools may also work with external partners, such as a research partnership or a school district/local authority or employ a 'critical friend' or external data coach to give support in the design of internal evaluation, the interpreting of evidence and to prompt reflection and planning for improvement. However, internal evaluation is distinguished from external types of evaluation in the fact that members of the school's professional personnel are in charge of the evaluation. External evaluations would, on the other hand, see an external authority (e.g. Inspectorate of Education) decide on evaluation criteria, collect data on school performance and report evaluation results.

'Data use' and internal evaluation

According to our definition, data- driven evaluation, based on test data only, is outside the scope of our review. In examining the literature it was found that the distinction between internal evaluation and data use, based on test data only, and driven by pressures for test- based accountability, was not entirely clear. 'Data- driven', 'data- centred', 'data use' or 'data- based decision making' (for example as used in Schildkamp and Visscher, 2013) were found often to describe processes which included the analysis of external test data, but which drew on other sources of evidence to inform collaborative discussion about ways to improve teaching and learning. In the USA, researchers found that teachers' understanding of the terms 'data' and 'data use' varied from school to school with only some schools employing the term in a way which is in scope according to the definition used for this review (Ikemoto and Marsh, 2007; Jimerson, 2014). Furthermore, within the range of studies of collaborative data use which involved more than test data, there was a range from quantitative data only, including assessment data and sometimes non- assessment data such as attendance and demographic data (Dembosky et al., 2006) to enquiry methods in which multiple sources of quantitative and qualitative data were discussed in professional learning communities (as, for example, in Farley- Ripple and Buttram, 2014). The selection of literature for this review aimed to be comprehensive, so that literature on data use is included, where it is clear that internal processes are part of the evaluation, where evidence other than external test data informs the evaluation and where additional insight is given into effects, conditions and mechanisms. It is acknowledged that the boundaries for inclusion may have excluded some relevant work.

2. Methodology

The review is focused on internal evaluation of <u>schools</u>, and excludes evaluation of individuals (e.g. head teachers, teachers). Work included was restricted to that published from 2000 until April 2015, from high income countries, in relevant settings (primary and secondary education) and language (English). Literature reviews on internal evaluation were used as a reference point for searches of relevant articles and books. Articles and books prior to 2000 were included if they were mentioned in these literature reviews and were relevant to the above four categories.

We looked at non empirical studies (e.g. think pieces, exploratory studies) to increase our understanding of the conditions and mechanisms that contribute to successful internal evaluation. In reporting findings we clearly distinguish which results are from empirical studies and which are from exploratory studies or think pieces. Publications for review were drawn from educational research, official government and international body publications, and a variety of other sources of evidence, including internet resources. Overall, the literature reviewers erred towards including studies where there was a degree of ambiguity regarding their relevance in order to aim for a comprehensive coverage of the field.

The literature review process:
This was carried out in steps and involved two research officers with general educational knowledge but non- expert in the field. The process was overseen by a Senior Lecturer who is an expert in the field of Inspection and evaluation. Refinements to the process of data extraction, understanding of key concepts and search terms took place through regular meetings and discussions.

Step 1
Searches: A research officer carried out an initial search to identify possible books, articles and reports on both inspection and on internal evaluation. Lists of titles of possible references were identified through searches of general databases, journal searches and searches of library catalogues. All keywords listed below were used in the searches of general databases and lists of titles were scanned for relevance to internal evaluation. This resulted in a large number of titles that were further filtered for relevance according to the above description of internal evaluation.

Journal searches and library catalogue search used more restricted sets of search terms to search 'keywords' and 'all text'. To add additional studies on internal evaluation, these were: 'school evaluation' and 'school accountability', 'data and school improvement'. An additional sweep of articles on evaluation was carried out by a researcher by referring to bibliographies and searches for authors frequently published in the field when the first set of searches on internal evaluation was found to have relatively few articles compared to those on inspection. Where titles were ambiguous, abstracts were referred to, if available, and more recent titles were prioritised. Manual scanning of database records was then used to refine title lists and to identify those authors who had published frequently in the field. As the steps of the search proceeded, additional titles were added, based on bibliographies of relevant titles and through use hand searches of journals and other relevant sources.

Step 2
Data extraction and summaries: Literature was filed and classified for inclusion, empirical studies were separated from non-empirical. Previous literature reviews were not included in counts but used to inform searches relevant to this review. Where empirical papers contained elements of separate data sets for each country, these were counted as discrete studies. Also, if individual papers referred more than once to one data set/study, these were aggregated in the data extraction and counted only once. Evidence tables include brief summaries of findings on internal evaluation effects and of important conditions and mechanisms. These can be found in the appendix

Databases searched and search terms:

Sources were identified through:

−   General databases (e.g. Digital Resource Archive (DERA), British Education Index (BEI); Australian Education Index (AEI); ERIC; Web of Science); internet search engines and gateways (e.g. Google Scholar); websites of inspectorates, education charities, policy 'thinktanks', conferences, unions.

−   Library catalogues, such as:

- American Educational Research Association
- Australian Centre for Economic Performance
- British Educational Research Association
- Bristol Institute of Public Affairs
- Business in the Community
- CfBT Education Trust
- Consortium on Chicago School Research
- Department for Education
- Education Scotland
- Education Sector
- Estyn
- FORUM: for promoting 3-19 comprehensive education
- Her Majesty's Inspectorate of Education
- London School of Economics Centre for Economic Performance
- National Education Policy Center
- Ofsted
- Research for Action.
- Royal Society of Arts and Manufactures
- Social Policy Association Conference
- US Department of Education
- Professional fora (e.g. Times Education Supplement, Guardian Professional,…)
- Manual keyword searching of journals:
- American Journal of Sociology
- British Journal of Educational Studies
- British Journal of the Sociology of Education
- British Educational Research Journal
- Cambridge Journal of Education
- Canadian Journal of Educational Administration and Policy
- Comparative Education
- Current Issues in Comparative Education
- Early Education and Development
- Education 3-13
- Education Inquiry
- Educational Action Research
- Educational Administration Quarterly
- Educational Assessment, Evaluation and Accountability
- Educational Management Administration & Leadership
- Educational Policy
- Education Policy Analysis Archives
- Educational Research Review
- Ethnography and Education
- European Education Research Journal
- European Journal of Education
- European Journal of Training and Development
- Improving Schools

- International Journal of Educational Management
- International Journal of Education Policy and Leadership
- International Journal of Qualitative Studies in Education
- Journal of Education Policy
- Journal of Educational Change
- Journal of School Choice
- London Review of Education
- Management in Education
- Quarterly Review of Comparative Education
- Research in Education
- Scandinavian Journal of Educational Research
- School Effectiveness and School Improvement
- School Leadership & Management
- Studies in Educational Evaluation
- Studies in the Cultural Politics of Education
- Teacher Development
- Scanning lists of references
- Contacts with those in the professional networks of the research team and those suggested by key informants.

Keywords for internal evaluation

Internal evaluation, Internal/external audit, Critical friend, Compliance, Teaching school networks/alliances, underperforming schools, special measures schools, failing schools, schools (requiring or needing) improvement, good schools, outstanding schools, satisfactory schools, unsatisfactory schools, (coasting or stagnating or stuck) schools, educational monitoring, database management systems, decision support systems, educational indicators, information management, information systems, information utilization, management information systems, management systems, performance information, performance factors, performance management, performance indicators, program monitoring, progress monitoring, school performance, progress reporting, recordkeeping, records, school data use, school data based, data and school improvement, school self-evaluation, SSE, self-assessment, student evaluation of teacher performance, teacher evaluation, total quality management, database management systems, school monitoring, EMIS, school performance data, monitoring systems, school governance, education governance, school boards, Governing education, school autonomy, school efficiency, national information systems, school marketisation, feedback and school, quality control, quality review, quality management, dynamic school improvement, institutional evaluation, school peer inspection/review, school self- inspection, school self- review, school self-regulation, Ofsted, Estyn, HMIE plus other names of inspectorates, accountability and gaming, educational accountability, standards based accountability.

The tables below provide an overview of the type of studies included in the review.

Table 1. Number of empirical and non-empirical studies

type of study	frequency		frequency
multi- country empirical	9	number of separate country studied empirical	33
single country empirical	73		73
total empirical	82		106
multi country non-empirical	3	number of separate country studied non-empirical	0
singe country non- empirical	15		15
total non- empirical	18		18

Table 2. Countries included in the review

countries non- empirical	frequency	countries empirical	frequency
Denmark	1	Belgium- Flanders	6
England	4	Canada	2
Ireland	1	Chile	1
Israel	1	Cyprus	2
New Zealand	2	Denmark	2
Scotland	2	England	12
The Netherlands	2	Germany	3
USA	2	Hong Kong	2
International	3	Iceland	2
		Ireland	4
		Italy	1
		Mauritius	1
		New Zealand	6
		The Netherlands	8
		Northern Ireland	2
		Norway	2
		South Africa	1
		USA	32
		Wales	2
		Europe	1

Table 3. Publication year of studies

year of publication non- empirical	frequency	year of publication empirical	frequency
2000	1	1998	1
2001	1	2000	1
2002		2002	2
2003		2003	4
2004	2	2004	2
2005	2	2005	4
2006		2006	5
2007	1	2007	8
2008	2	2008	10
2009	1	2009	3
2010		2010	6
2011		2011	7
2012	2	2012	7
2013	4	2013	6
2014	2	2014	13
2015		2015	2
	18		81

3. Findings

This chapter presents the findings from the review. We will first present a more elaborate definition and description of different types of internal evaluation. We then go on to present our findings on different types of effects and side effects of internal evaluation, while the last two sections explain the conditions and mechanisms of effective internal evaluation.

3.1 Different types of internal evaluation

Internal evaluation involves a number of stages:

Plan – school staff and other relevant stakeholders meet to discuss and agree the aims of the evaluation.
Do – collect data for the purposes of evaluation, such as student assessment data, demographic data, student work, lesson observations, student, staff, parent questionnaires or interviews and any other data deemed relevant to the evaluation.
Check – analyse the data collected and use this to assess the extent to which aims are being met and where learning needs of students are not being met. This stage may identify professional development needs for teachers and leaders.
Act – Follow up findings and recommendations from the review process.

Internal evaluations are formative when focused on assessing strengths and weaknesses to inform continuous improvement and school development or may have a summative function when internal evaluation reports (from the *Check* phase) are used for external accountability, for example when they feed into external evaluation such as inspection. Not all jurisdictions have an external quality assurance system for schools. In Cyprus, for example, it is individual teachers who may be subject to inspection. Nevertheless, schools may use internal evaluation as a *stand alone evaluation* to inform school improvement.

In jurisdictions where there is a system for external quality assurance, this may interact with internal evaluation or be independent of it. In *sequential evaluation,* schools conduct their own evaluation and the external body (for example, the Inspectorate) then uses the internal evaluation as a basis for its external evaluation (Kyriakides and Campbell, 2004). The external evaluator may analyse the internal evaluation data and it may then be validated, or otherwise, by data collected by the external evaluator. This is the case, for example, in Hong Kong (Wong and Li, 2010). Alternatively, the external evaluator may use the data from the internal evaluator to contribute to a single meta- evaluation. According to Kyriakides and Campbell (2004) sequential evaluation may also work in the opposite direction, when the external body provides feedback to the school which is expected to be used in internal evaluations and improvement of the school. In this case, the action plan for improvement of the school will reflect the criteria and judgements of the external evaluator instead of school- defined criteria. These variants on the sequential evaluation model are not necessarily clearly distinguishable in the literature. For example, in England, inspection reports identify 'areas for improvement' which are expected to form the basis for the school's action planning and own internal accountability processes. Internal evaluation, or school self- evaluation (SSE) is expected in England and, for a period prior to 2010 was mandatory, using a specified tool the self- evaluation framework (SEF). When the school was inspected, inspectors viewed the SEF as part of their preparation for the inspection. According to Davies and Rudd (2001), there was a conflict for staff in schools between preparing for Ofsted and conducting the internal evaluation, with teachers not sure how much time to dedicate to each activity, which were seen as separate.

Janssens and Van Amelsvoort (2008) studied internal accountability in seven European countries: England; Scotland; the Netherlands; Northern Ireland; Denmark, Belgium and Germany (Hesse and Lower Saxony). The study explores the extent to which each system orients towards accountability (Accountability Orientation or AO) or improvement (Improvement Orientation or IO). All countries

used a form of sequential evaluation in which internal evaluation was used to prepare for inspection to some degree and provided frameworks on school improvement to guide internal evaluation. In Denmark and Hesse, the role of the inspectorate is to support and provide advice and the frameworks here were less standardised and allow for more freedom for the school to pursue its own priorities for improvement, as was also the case in Belgium (improvement orientation). In countries with a strong accountability orientation, where the internal evaluation contributed to inspectors' decisions (England, Scotland, Northern Ireland, Lower Saxony) the authors suggest that there is potential for conflict between using internal evaluation for school improvement and an external accountability function.

A cooperative model (Kyriakides and Campbell, 2004) or *collaborative evaluation* (Christie et al., 2004) integrates internal and external evaluations in the first three stages; *Plan*, *Do*, *Check* and internal and external evaluators collaborate to plan, design, conduct and report on the evaluation together. As a result, the interests and viewpoints of external evaluators, as well as the school- defined criteria are taken into account simultaneously. According to Kyriakides and Campbell (2004) measurement criteria come up from both types of evaluation and an attempt is made to combine the results in order to conduct a holistic evaluation and satisfy the needs of both parties.

In *parallel evaluation*, internal and external evaluators do not participate in each other's evaluation. According to Kyriakides and Campbell (2004) and Christie et al. (2004) the school and the external body conduct their own evaluations, which may be shared when completed. This is the case in systems such as in much of the USA, where external evaluation or accountability is test- based, with student performance in externally set tests used to make performance judgements on schools and teachers.

3.2 Effects and side effects of internal evaluations

There is sufficient evidence from international research to show that internal evaluation *can* lead to sustainable school improvement with increased student achievement and better teaching and learning, although, as is shown in later sections, findings on the positive effects of internal evaluation are frequently qualified by the need for suitable conditions and mechanisms to be in place. Effects are discussed in relation to: reflection on school quality and intentions to improve; school improvement; improved student achievement and improved conditions of learning. Examples where internal evaluation has led to no effect are also discussed.

3.2.1 Reflection on school quality and intentions to improve

Although internal evaluation results may not lead directly to improvement actions they may influence thinking and the direction of school improvement (Schildkamp et al., 2012). Nevo (2001) suggests that internal evaluation can lead to greater sensitivity to areas in need of improvement through what Simons (2013) calls "insights into the educational experiences of students" (p 2). Schildkamp et al.'s (Ibid.) study compared the effects of systems for internal evaluation in the Netherlands and in Flanders. In the Netherlands, internal evaluation was found to lead to more frequent and open consultation about the quality of education and more classroom visits by the school leader. Davies and Rudd (2001) carried out research in England at a time when internal evaluation was at an early stage in many of the 23 schools studied. They found that the process of internal evaluation led to an increased use of classroom observation and involvement of parents and pupils. It helped to identify professional development needs and led to greater ownership of change. Use of relevant, school- specific data was found to be important in developing critical reflection and evaluation in a study in Northern Ireland, in which teachers were asked to keep a research journal over a period of four weeks (Neil and Johnston, 2005). The process of internal evaluation extended existing methods of evaluating teaching and learning, identified priorities for the professional development of teachers and developed a perspective among teachers beyond their own classroom. In Iceland, Davidsdottir and Lisi (2007) found changes in teacher perceptions of administration facilitating their professional growth; systematic data collection; teacher ownership and collaboration; and shared decision-making for improvement efforts in the schools.

Questionnaire responses from teachers in Denmark, England, Finland, Scotland and Sweden showed that in all countries, teachers believed that internal methods of quality assurance, including internal evaluation, analysis and tracking of pupil progress and school target setting had the most influence on improving their practice (Grey et al., 2011). This data was collected as part the European Fabricating Quality in Education project and contrasted with many teachers' beliefs that external aspects of quality assurance, including external tests, comparison with other schools, publication of data and inspection did not contribute to improvements to such an extent. In Denmark, for example, 85% agreed that internal quality assurance of schools in more beneficial than external methods.

In systems where an inspection system is in place, sequential internal evaluation helps schools prepare so that they know where they are in relation to inspection standards (Davies and Rudd, 2001; Ritchie, 2002; Yeung, 2011). In systems where there is external verification by an inspection process, this further helps to develop a more rigorous and systematic way of assessing practice and the capacity of the school to improve (Ehren et al., 2014; MacBeath, 2008).

Internal evaluation can also be of value in systems without external criteria or verification set through inspection in a stand- alone system. In Iceland and Ireland respectively, McNamara et al. (2011) and Karagiorgi et al. (2015) provide examples of small groups of primary schools, which each developed their own process and indicators with the support of external facilitators to embed a culture of self-evaluation, reflection and improvement. In Ireland, a self- evaluation process supported by an external facilitator built capacity by raising expectations and helping the staff in small post- 16 units become more professional (O'Brien et al., 2015).

3.2.2 Effect on school improvement

Improved critical reflection based on internal evaluation can inform goals and actions for improvement planning (Bubb and Earley, 2008; Caputo and Rastelli, 2014; Cowan, 2008; Demie, 2003; Karagiorgi et al., 2015; Supovitz and Klein, 2003; Wohlstetter et al., 2008). Ehren et al. (2014) suggest that where there are inspection frameworks that set expectations about standards in education, these can drive internal evaluation and the capacity of a school to improve. Cowan (ibid.), in an evaluation of the internal evaluation (school self- evaluation) policy then current in England, based on survey data collected over two years found that improved data availability supported improved self – evaluation, which in turn supported an increased focus on achieving outcomes for pupils overall and for specific groups, a better understanding of underperformance, better engagement with students and parents and increased accountability among middle leaders. Supovitz and Klein (ibid.) used case study and survey data from schools in the USA that were part of the *America's Choice* reform programme, which included systematic and school wide use of data. They found that schools used the data for school improvement through planning of professional development, setting goals, motivating teachers and students, visually stating school priorities and goals and communicating with parents. Hall and Noyes (2007) found that in the collaborative schools identified in their study in England, internal evaluation was seen as contributing to school improvement as a shared endeavour involving middle leaders and teachers.

Implementation of a range of improvement strategies based on priorities identified through internal evaluation can lead to improvement across a school (Education Scotland, 2012; Estyn, 2012; Neil and Johnston, 2005; Schildkamp et al., 2012; Supovitz and Klein, 2003; Leung, 2005). Thus, in Wales, the chief inspector's report for 2012 (Estyn, 2012) found that schools had increased professional learning, revised content or organisation of the curriculum and provided targeted support for groups of pupils.

3.2.3 Effect on student achievement

In Caputo and Rastelli's (2014) study of the impact of an internal evaluation training programme for schools in Southern Italy, the quality of school improvement planning was found to influence the effect on student achievement that resulted. They used a standardised mathematics test for students at the beginning and end of a year- long support programme, in which teacher training on internal evaluation

and action planning was supported by opportunities for working in professional learning communities. Analysis of action plans showed that those schools which had produced a more accurate evaluation of the school context and more specific improvement goals had higher student achievement gains than those schools which had been less specific in relation to improvement. Demetriou and Kyriakides (2012) worked with sixty primary schools in Cyprus, which they divided into four groups, one of which was a control group. All of the remaining three groups conducted an internal evaluation and planned for improvements, with one of these groups receiving support based on a theoretical framework informed by school effectiveness principles. The impact of internal evaluation on student achievement was measured through performance in a mathematics test at the beginning and end of the year- long project. The school which conducted internal evaluation using the framework informed by school effectiveness principles saw significantly greater gains than all the other groups. All of the groups of schools that had conducted an internal evaluation showed better improvement in student achievement than in the control group.

Cowan (2008) found that school self- evaluation in England over two years had led to improved standards in many schools, particularly for groups that the evaluation process had identified as under-achieving. Studies of schools where there was systematic and supported use of data found significantly increased achievement in the USA (Cosner, 2011; Gallimore et al., 2009; Marsh et al., 2010), Canada (Dunn et al., 2013) and Wales, (Stringfield et al., 2008). In New Zealand Lai and McNaughton (2013) discuss a model with three phases: initial discussion and analysis of data; professional development targeted to identified needs; professional learning communities which aimed to integrate changes into the normal routines of the school. Groups of between 4 and 34 schools worked in clusters, with each supported by a researcher. Student progress was greater than expected, with cumulative gains over three years. The method was replicated and reported in a further study (McNaughton et al., 2012) with similar results. Also in New Zealand, Timperley and Parr (2009) found that professional development for leaders and teachers based on needs identified through a whole- school assessment which included classroom observation and a questionnaire, resulted in student achievement gains greater than expected, particularly for low- attaining students.

3.2.4 Effect on conditions of learning

In the Timperley and Parr (2009) study, noted above, as well as recording gains in student achievement, classroom observations at the beginning and end of the project showed evidence of considerable change, with changes in instructional practice and teachers' pedagogic content knowledge confirmed in interview evidence from leaders and teachers in the project schools. In McNaughton et al.'s study (2012) it was the development and use of context- specific teaching programmes that was found to be effective in raising student achievement. Similarly in the USA, Gallimore et al. (2009) studied 15 schools over a period of five years, with nine using an experimental inquiry- based protocol for internal evaluation and six comparison schools using other school improvement models. They found that schools using an enquiry- based protocol to examine student data fostered the acquisition of teaching skills and knowledge and instructional planning to significantly increase achievement over than in comparator schools. Rather than attributing student achievement to external causes, teachers shifted over the course of the project to examine their own teaching for ways to improve achievement. Farrell (2014), also in the USA, in a study of secondary schools' data use, points to specific changes in practice such as changes to vocabulary in lessons, re- teaching where students had not understood a concept and targeting specific students. Wayman and Stringfield (2006) found that in three schools in the USA that used data effectively there was an increased sense of teacher efficiency, a better response to student needs and more collaborative reflection on practice and Halverson et al. (2007) point to the ongoing dialogue between teachers and leaders on the use and implications of data, which led to improvement in instruction. Improvements in teaching following internal evaluation were also reported elsewhere in the USA (Copland, 2003; Dembosky et al., 2006; Marsh et al.,2010), Germany (Hartong, 2012), the Netherlands (Hofman et al., 2010). In Hong Kong, MacBeath (2008) found that teaching was more engaging and learner centred.

Consulting with pupils as part of the internal evaluation, contributed to improvements in practice in England (Wroe and Halsall, 2001) and Finland (Webb et al., 1998), with supportive peer observations also reported as leading to improvements in England (Chapman, 2000).

3.2.5 No effect

In contexts in which internal evaluation is not established nor formally supported, as in Mauritius (Ah Teck and Starr, 2014), Cyprus (Karagiorgi, 2012) and Uruguay (Vazquez and Gairin, 2014) little change resulted either from the informal methods reported by Ah Teck and Starr or following the use of an optional tool by Vazquez and Gairin. In Karagiorgi's (2012) report on a project in Cyprus, teachers in a small primary school liked the collaborative approach taken to the introduction of a school self-evaluation and found it easy to identify a priority for improvement. However the planned intervention was not implemented with teachers claiming to lack the time and resources to make changes.

In systems with more widespread use of internal evaluation, the type of support provided may be significant. Timperley and Parr (2009) reported on the comparative impact of two interventions intended to support schools in using data for instructional improvement in literacy. In one study, cited above, there were improvements in teaching and improved student progress. In the other intervention, training was provided for school leaders in using data to inform practice, but there was no school- based needs analysis and identification of teachers' and leaders' learning needs or professional development tailored to meet these needs. No impact was shown on student achievement in the study which was limited to training of school leaders and the researchers found that there were different understandings by policy makers and school leaders. A project in the USA described by Quint et al. (2008), which put data coaches into schools to work with teachers on identifying implications for instruction from student data found no difference between project schools in the project and comparator schools. Project schools commented favourably on the data coach role and the professional development they provided but comparator schools also reported similar levels of professional development on data use and spent as much time analysing data. The authors note that because both sets of schools were in a district in which use of data to inform teaching was encouraged overall, their study is insufficient to judge the significance of a data coach role. In the Netherlands, Blok et al. (2008) considered a model in which school's own internal evaluation was followed by validation from external visitation and inspection. Internal evaluations were found to be low quality, often failing to answer questions set at the onset of the process. Blok et al. conclude that considerable support and guidance was needed for schools in completing internal evaluations and at the time of the research (2003-6) their completion was unlikely to lead to school improvement. Although the report of the chief inspector for Wales (Estyn, 2012) comments favourable on the impact of data use on school improvement in most inspected schools in Wales, they found that about one- fifth of schools were not using data effectively to plan for improvement and recommend that training and support is provided for these schools. In Scotland similarly, the chief inspector's report (Education Scotland, 2012) notes that the use of self- evaluation to drive improvements could be more effective and "approaches to self- evaluation need to impact on young people's learning and achievements, including attainment".

3.2.6 Unintended effects

Andersen et al (2009) discuss what they term 'measure fixation' in the context of Denmark's quality assurance system which includes both internal and external evaluation. Measure fixation happens when practitioners focus on exactly what is being measured as an indicator of quality, often at the expense of genuine quality. Under advanced measure fixation, the indicator provides a definition of quality along with an indicator of how to measure quality, so that it is not possible to distinguish between genuine quality and quality measured by an indicator, since the latter helps define the former. Hartong expresses similar concern in the research on use in Lower Saxony, Germany on a school self- evaluation tool (SEIS) developed by the Bertelsmann Foundation. The tool collects data from teacher, student, parent and principal questionnaires and provides a report back to the school, and, she claims "Despite critique on the SEIS criteria or framework, the correctness of the evaluation results is invariably accepted" (p 755) and it is used for action plans. Improvement is measured through subsequent SEIS results, such

that what is good practice and the measurement of this are derived from the same source, with loss of school and teacher autonomy in determining what they mean by quality.

A framework, or prescribed model for internal evaluation, may be helpful to schools but may also have negative effects. Croxford et al. (2009) describe expectations for internal evaluation in Scotland and the role of the local education authority in supporting this. They questions the extent to which a top down approach, in which performance goals are set by the inspectorate can be compatible with the school improvement goals of teachers. The authors suggest that the model may encourage performativity and compliance rather than improvement. Based on her experience internationally of internal evaluation, Simons (2013) reports that teacher resistance may prevent top- down initiatives from being successful, especially where unions are strong, as in Ireland, or teachers' motivation is weak, as in Spain where teachers are civil servants for life and there is no institutional leadership structure to drive change. She adds that test- based external accountability may cause schools to neglect other achievements and priorities for evaluation.

Davies and Rudd (2001) and Hall and Noyes (2007) in England, Wong and Li,(2010) in Hong Kong, Vazquez and Gairin et al. (2014) in Uruguay note the stress and heavy workload associated with internal evaluation. MacBeath (2008) also reporting on the system in Hong Kong, where internal evaluation is validated by external review, noted the heavy workload together with high levels of stress and anxiety among teachers. He comments that this can be managed by school principals, who can lessen stress by emphasising the learning opportunity that is offered by the process.

Although data availability can be helpful in internal evaluation, Wroe and Halsall (2001) note that too much, or the wrong kind of data, is unmanageable and confusing.

3.3 Conditions for effective internal evaluation

The majority of the literature studied considered the *process* of internal evaluation, with only some discussing outcomes. All authors agree that in order for internal evaluation to be successful in objectively examining the school context and the learning of students and in using the findings to prompt discussion about improvement, a number of conditions must be in place. These may be summarised as: evaluation literacy, resources, leadership, external support, supportive climate and accountability.

3.3.1 Evaluation literacy

Earley and Bubb (2014) emphasise the need for research- literate staff in schools and quote the British Educational Research Association (BERA) who say "Teachers need […] to be equipped to engage in enquiry-oriented practice – having the capacity, motivation and opportunity to use research-related skills to investigate what is working well and what isn't fully effective in their own practice. Using data from a number of different sources teachers can identify problems together with interventions which are likely to lead to improvements. (p 30)" (BERA, 2014). Timperley (2013) agrees that "the capacity of schools to reflect on the quality and accuracy of their data and to perform accurate analyses relevant to their purpose is widely viewed as an integral part of effective self- review and evaluation" (p63). Expertise needs to be spread throughout the school (Geijsel et al., 2010). There is a need for expertise in drawing on several forms of evidence and mediating differing understandings to identify learning and improvement needs (Chapman, 2000; Coburn and Talbert, 2006; Cosner, 2011; Lai and McNaughton, 2013; Stringfield et al.,2008; Wayman et al., 2007) as well as for technical expertise in analysing student assessment data (Supovitz and Klein, 2003).

The extent to which there was steering for schools through a framework and expectations about the process for internal evaluation were found to be important features in international research by Janssens and Van Amelsvoort (2008). Based on research in Ireland, McNamara and O'Hara (2008, 2012) say that such expectations and frameworks for internal evaluation are insufficient and that schools need

support in developing necessary skills. In focus group discussions in 2011, school leaders said that guidelines provided on the evidence base for internal evaluation were not clear and there was a lack of benchmarking data for student attainment. The need for support in developing the skills and building capacity to evaluate and to interpret data are further emphasised across the international literature (Anderson, 2010; Blok,2008; Caputo and Rastelli, 2014; Estyn, 2012; Honig and Ikemoto, 2008; Jimerson, 2014; Lachat and Smith, 2005; Marsh et al., 2010; Murnane et al., 2005; Mutch, 2012; Schildkamp and Visscher, 2013; Schildkamp, Lai and Earl, 2013; Schildkamp et al., 2014; Simons, 2013). Lai and Hsiao (2014) found that, even with support, only two- thirds of the groups of schools studied could produce high quality self- evaluation data.

Without the skills for evaluation literacy, there is a danger that schools may treat the process as superficial, as in the Irish schools studied by McNamara and O'Hara (2012) and the 'resisting' schools in England (Hall and Noyes, 2007) where it was often seen to be a compliance exercise rather than genuinely aimed at school improvement. Barrett (2009), Little and Curry (2008) and Timperley (2008) looked at the detail of teachers' collaborative conversations about data. They all characterize the talk as frequently superficial and lacking purpose, with Barrett adding that it focused on explanations for student failure that reside outside the control of the teacher, rather than looking for how teaching can be improved to meet students' needs. In her study, Barrett found that the presence of a facilitator and tools for displaying and reviewing data appeared to have little effect on what seemed to be ingrained ways of discussing students according to perceived effort, motivation and ability. Lasky et al. (2008) also point to the need for skills that allow for intentional and possibly critical conversations anchored to student data that can inform teaching and school improvement plans. Chapman (2000) similarly discusses the need for expertise to identify student needs, including taking their views into account (Hofman et al., 2010).

Validation of internal evaluation to add challenge and rigour has been suggested (Chapman, 2000; Trachtman, 2007) with both authors suggesting that partnerships with other schools may support this. Trachtman, writing of the USA, where there is no formal framework or standards to set expectations, suggests school partnerships or partnerships with higher education institutions may provide a form of peer review that can help validate judgements about quality against agreed standards.

3.3.2 Resources

The most commonly mentioned resource need is that of time, with time for developing and embedding evaluation literacy mentioned as a condition for successful internal evaluation (Copland, 2003; Cosner, 2011; Davies and Rudd, 2001; Dembosky et al., 2006; Farrell, 2014; Marsh and Farrell, 2015; Ryan et al., 2007; Schildkamp et al., 2014; Supovitz and Klein, 2003; Wayman and Stringfield, 2006; Wohlstetter et al., 2008). Wohstetter et al. found that it took three years in one of the school districts studied for all teachers to be able to articulate goals clearly. The importance of the provision of dedicated time for collaborative discussion as a key element of internal evaluation is specifically mentioned by several authors (Copland, 2003; Kallemeyn, 2014; Lachat and Smith, 2005; Means et al., 2010; Schildkamp et al., 2012). Time is also mentioned in regard to implementing improvements which have been planned as a result of internal evaluation. Marsh and Farrell, (2015) found that lack of time prevented response to some identified needs and this was also one of the reasons that teachers gave for not introducing planned interventions following an internal evaluation process in a school in Cyprus (Karagiorgi, 2012).

Data systems, data availability and timeliness of this is a further resource need identified (Cowan,2008; Dembosky et al., 2006; Farrell, 2014; Lai and Hsiao, 2014; Schildkamp et al., 2014; Wayman and Stringfield, 2006; Wayman et al., 2007). The availability and sufficiency of training, specifically in the use of data and data systems, but also more broadly in relation to evaluation literacy was a further resource gap (Dembosky et al., 2006; Ryan et al., 2007)

Tools to support internal evaluation may help schools (Davies and Rudd, 2001; Farrell, 2014; Leung, 2005) so long as suitable training is available and provided (Dembosky et al. .2006; Farrell, 2014; Schildkamp et al., 2014). Verhaege et al. (2013) examined five such tools to support internal evaluation and discussed their use with advisory staff in New Zealand, England, South Africa, the Netherlands and Belgium (Flanders). They argue that schools should be aware of a tool's purposes and mechanisms and of the ownership of any data generated before electing to use a particular supportive tool. Hofmann et al. (2005), based on examination of available tools in the Netherlands, proposed a framework for comparing internal evaluation tools according to an accountability objective, a school improvement objective and for reliability and validity and also argued that schools should use the framework to guide choice of a suitable tool. Both Hartong (2012) and Wohlstetter et al., (2008) express concern about the potentially restrictive consequences of system- wide frameworks for supporting evaluation and improvement, with both commenting on the prescriptive models for curriculum and teaching recommended as improvement measures. Hartong suggests that what is recognised as 'good' becomes more limited as a result. Wohlstetter et al. suggest that standard, uniform approaches to curriculum and teaching are positive for new and poorly- performing teachers, but that these are restrictive for higher-performing teachers.

Lack of resources is also mentioned as a barrier when internal evaluation has not been successful, with Farrell (2014) reporting that the resources provided influenced the extent to which data was used to support improvements in teaching.

3.3.3 Leadership

The importance of effective leadership at all levels recurs in the literature on internal evaluation. Leadership activity may be directly in support of internal evaluation, for example through modelling data use or in leading collaborative discussions, or indirect in ensuring that resources are provided and that a culture is developed which is improvement oriented and enabling of critical reflection and challenge to existing practices. As Earley and Bubb (2014) comment, leadership is required to create and embed an inquiry- oriented culture and to provide the resources, particularly of time, to allow staff to collaborate, agree common goals for inquiry and discuss findings and to put support in place for staff to develop research expertise. Implementing improvements to impact on teaching quality and student outcomes may also require strong leadership (Bubb and Earley, 2008; Devos and Verhoeven, 2003). Although leadership by the school principal is most frequently mentioned, a need for distribution of leadership throughout the school is often emphasised. In the local authority or school district effective leadership is a strong influence on school practices in internal evaluation.

Ah Teck and Starr (2014), in a system without an established system of internal evaluation found that principals were resistant to any formalised quality assurance process. In Karagiorgi's (2015) case study of introducing self- evaluation in a school in Cyprus the support of the principal was considered essential. For Demie, (2003) who described the introduction of a data system in a local authority in England, the leadership of the headteacher was essential to raise expectations and, in Norway, Emstad (2011) says that the principal has a key responsibility in both prioritising internal evaluation and ensuring the time for discussion. In the Netherlands, where internal evaluation is expected, its quality is still dependent on the attitude of the principal, according to research by Geijsel et al., 2010, who found that principals need to share an aim of learning from data and understand that the intention is not to judge the school from the outcome, but use it to make changes. Principals need to convince teachers too that data is used for diagnostic purposes that are beneficial and non- threatening (Wayman and Stringfield, 2006). Leadership committed to data use for improvement is vital for Marsh and Farrell (2015); Lachat and Smith (2005); Murnane (2005); Neil and Johnston, (2005); Schildkamp et al., (2012); Schildkamp and Visscher, (2013); Schildkamp et al. (2014). Schildkamp and Visscher's (2013) article is based on experience in two projects in the Netherlands, with the first, the FOCUS project, involving 150 primary schools in an approach which provided training and support for schools in an improvement cycle that uses internal evaluation of student monitoring data as a starting point for setting challenging goals and designing and implementing a learning strategy to address these. Their list of critical conditions for successful internal evaluation are all such as to be dependent on a school leader

who stimulates and facilitates the use of data: motivated staff; a school culture that is achievement-oriented; collaborative agreement on a clear set of goals and clear division of tasks between team members; a cycle of core activities; knowledge and skills for data use; pedagogical content knowledge. Not all school principals are able to recognise and act on their responsibility to secure necessary conditions for internal evaluation. In Anderson et al.'s (2010) study, principals were often aware of conditions that were required to foster or inhibit data use, but only a few took actions to shape the conditions, such as developing teachers' evaluation literacy and through provision of time, that fostered use of data for improvement in their own schools.

Several authors who focus on the role of the school principal, or headteacher, note the way in which this changes over time, as internal evaluation is first introduced into a school, plans are developed and implemented and the process is repeated in a continuous improvement cycle. Copland (2003) describes the role of the principal or other formal leaders as catalysts in initial stages, with distributed structures emerging as the model develops. Principals maintain the vision for implementing improvements by firing and hiring staff and through providing time for collaboration. Cosner (2011) similarly describes principals' contribution to the design and introduction of tools and processes that support data- based collaboration in schools. As a state- wide system for data use was introduced in Ontario, Canada, principals were expected to promote collaborative discussion and learning from data in professional learning communities and for monitoring changes in classroom practice (Dunn et al., 2013). Young (2006) found that where leaders did not provide vision, expectations and norms for collaborative discussion of data to improve instruction across teams, teachers worked individually without collaboration.

Some studies refer to the principal being directly involved in steering discussion on the implications of data collected for learning in the school. Earl (2008) refers to leadership of collaborative discussion in a case where the principal used data charts to maintain a focus on discussing implications for practice in a group where teachers tended to deviate into more general discussion of teaching or factors that affect learning. Kallemeyn (2014) describes how the principal directed and modelled a school- wide cycle for improvement in a case study of an elementary school in the USA. The school principal allocated time for team to meet and set clear expectations for use of this time to review data with and instructional coach and use it to plan for teaching and learning. Whole- school learning walks were used, where grade teams were encouraged to share the progress of students with one another.

The importance of the school district and the principal in establishing the vision for internal evaluation and being engaged in the process is highlighted by comparisons made by Farley- Ripple and Buttram (2014) who compared developments in the US state of Delaware, following a state- wide allocation of time for teachers to meet in professional learning communities to discuss data and use it for improving teaching and learning. Clear differences emerged between schools in different districts in the state. Where leadership was absent, with a lack of drive and support from both district and school leaders, there was less evidence for the impact of collaborative data use that in those where they were active.

In systems where there is high external accountability, the vision for internal evaluation established by the principal must balance external requirements with their own beliefs and knowledge of their school context. The principals studied by Knapp and Feldman (2012) were able to do this by identifying commonalities to inform a vision for education in their school which teachers could share, with collaborative working encouraged. Information provided by both internal and external evaluation was integrated to inform professional development, with modelling from the principal on how the information might be used to make improvements. Ehren et al. (2014 identify the principal as the mediator between the school and its context and the demands for external accountability, with their response determining policy and structures within the school.

Although the leadership of the principal is considered essential for successful internal evaluation, ownership of the process through distribution of leadership is equally important (Leung, 2005; Vanhoof

and Van Petegem, 2011). Means et al. (2010) found leadership of data use extended beyond the principal to individuals in a variety of job roles, such as instructional coaches and department lead teachers.

3.3.4 External support

Internal supports for internal evaluation, particularly those studies using terminology of data use, include roles such as data coaches or lead teachers who are employees of the school (Farrell, 2014; Gallimore et al., 2009). External support provided by a partnership with a university or group of other schools, through the individualised support of one or more researchers is often mentioned as crucial in developing evaluation literacy within schools and within groups of teachers. This may be through supporting inquiry methods or discussions in professional learning communities. Another common form of external support is that provided by the objective viewpoint of a person who is not a member of the school staff and who may challenge the school as a critical friend. In some cases, the challenge is linked to accountability expectations. Supovitz and Weathers (2004) describe a system in which objectivity and challenge are provided by peer review in a school district, with overall learning shared in principals' workshops.

Support from an external source in developing evaluation literacy for internal evaluation literacy may include the provision of training, as in the model for data use described by Demie (2003) in a local authority in England. Janssens and Van Amelsvoort (2008) in a multi-national study in Europe found that having access to a range of training providers, not just from the inspections system, was beneficial to the improvement orientation of the evaluation. When Ontario, Canada introduced a data use system, these were supported by workshops and professional learning communities, supported by a trained facilitator, within schools (Dunn et al., 2013). In Germany, the use of the SEIS tool, developed by the Bertelsmann Foundation to collect and analyse data from student, teacher, parent and principal questionnaires, is supported by the provision of consultants and coaches (Hartong, 2012). External facilitators working with groups of teachers from one or more schools is considered essential by some researchers, helping schools by providing professional development, access to research on learning, facilitating collaborative discussion and developing tools (for example for learning walks in Honig and Ikemoto's study) to help teachers (Honig and Ikemoto, 2008; Karagiorgi, 2015; Lai and McNaughton, 2013; McNamara et al., 2011; McNaughton, Lai and Hsiao, 2012; O'Brien et al., 2015).

Some examples in the literature describe a partnership between researchers, based in a university, working alongside schools on internal evaluation. A benefit of such a partnership is that researchers are able to be responsive to emerging goals and needs in the schools with which they work (Ancess et al., 2007; Davidsdottir and Lisi, 2007; Hermann and Gribbons, 2001; Sjobakken and Dobson, 2013). To illustrate their point, Ancess and his colleagues present two case studies. In the first case a researcher from the university worked with a consortium of three schools that were considered to be innovative in their approach and which wanted to use data to help improve mathematics learning. The researcher helped by providing a structure and support for a working group from the three schools, helping them to work through the analysis of data and the planning and implementation of improvements. In the next cycle of data use, learning was used to revise the structure and provide additional resources and data, so that a cycle of inquiry developed in the group of schools. Challenges of different levels of expertise and confidence among the working group needed to be overcome and the researcher needed to keep discussion focused on the learning from data. In the second case, the challenges were more difficult, such that the researcher initially worked with a representative of the school staff, the data coach, to plan for a session on sharing students' work as the least threatening to teachers who were used to working in isolation and were wary about sharing practice and visiting each other's classrooms. Trust was then built gradually, with the researcher providing resources to fit the school's context. Teachers began to see the internal evaluation process as something in which they had ownership, rather than being 'done to' as recipients. In Sjobakken and Dobson's example, a researcher worked with a Norwegian school over a period of eight years in developing a process of internal evaluation to support improvements in education for students with special educational needs. Methods for collecting data, such as examination of pupil work, video recordings and reflective 'letters to the researcher' were negotiated between the

researcher and the school. Sjobakken and Dobson suggest that the process of self- evaluation is an ongoing process, where negotiated ownership developed over time is needed to secure sustainability.

Swaffield (2004) describes a critical friend as one who can offer a different perspective, ask provocative questions and act as a sounding board (p275) to support internal evaluation and the value of this role is endorsed in other work from England (Davies and Rudd, 2001; Cowan, 2008). The objective, external perspective provided by researchers working with three schools in Flanders helped to identify blindspots for those internal to the school (Devos and Verhoeven, 2003). However, Meuret and Morlaix (2003) reported that the impact of the critical friend had been weak, despite being appreciated by respondents.

In a later publication, Swaffield gives an alternative definition of the role of critical friend as a moderator, in confirming the findings of internal evaluation (Swaffield and MacBeath, 2005). This is more like the ways in which district officers influence the actions of school principals with regard to data use in the study by Anderson et al. (2010) or the monitoring of changes in classroom practice with the principal by a critical friend in Ontario (Dunn et al., 2013). Hofman et al. (2010) emphasise the importance of a critical friend as moderator, and state that key factors for effective internal evaluation are that independent external validation is sought for internal judgements on quality and that information about quality is shared with teachers, students and parents. Supovitz and Weathers (2004) describe a system in which schools' implementation of district policies is monitored through periodic visits to schools by a small team of district officers and peer principals, who collect data from the school and discuss their findings with leaders of the school. Across the district, overall findings of progress and issues are discussed in regular principals' workshops. The system is non- judgemental and provides constructive feedback to help schools amend strategies for improvement, while sharing learning across the district and developing a shared language for professional conversations.

3.3.5 Supportive culture

In discussing the role of the school leaders, the contribution made to the creation and embedding of a culture of enquiry, with commitment to the use of internal evaluation for improving the school, was found to be a necessary condition. Two aspects of such a culture are brought out strongly in findings from the research; the importance of trust and an orientation towards improvement.

A climate of trust in the school, underpinned by supportive relationships and good communications, allows for collaborative working in which teachers can be critical in their analysis of data and challenge one another openly. Timperley (2008) found that in the most improved school of the three she studied, teachers were more ready to express uncertainty in group discussions and to seek help from others in the group. Such a climate enables distributed leadership and shared ownership to develop in internal evaluation (MacBeath, 2008; Montecinos et al, 2014). The importance of trust, the absence of blame and a history of working together is emphasised by Earley and Bubb (2014), Marsh and Farrell (2015, Leung (2005) ,Vanhoof et al. (2012) and Wohlstetter (2008). Marsh and Farrell studied the impact of capacity building interventions in six, low- income schools in the USA that had all failed to meet state accountability standards in each of the previous five years, and in which a coach worked with teams and individual teachers. The researchers found that where trust and a history of working together was established, with values and expectations that supported open, critical inquiry around data and instruction, data use to improve instruction improved more rapidly.

In findings from a research project which covered several European nations, Schildkamp et al. (2014) found that where teacher collaboration was common, as in England, Lithania and Poland, interviewees were able to provide more concrete examples of the impact of decisions made based on the data reviewed in various groupings, such as subject department meetings. Wohlstetter et al. (2008) found that teachers relied upon one another for discussions about data, for support and for new instructional strategies. Trust is particularly important in peer observations of teaching (Chapman, 2000) so that feedback can be provided objectively and received positively.

Internal evaluation is more likely than external inspection to allow for trust and openness, as pointed out by McNamara and O'Hara (2006) where the lack of criticism from outside makes it easier for teachers to challenge each other and use the outcomes constructively. Where supportive relationships and trust were not in place, isolation and hierarchical leadership were barriers to change (Copland, 2003) and to the open dialogue needed for evaluation (Ryan et al., 2007). As Herman and Gribbons (2001) point out, the importance of trust and what they term efficacy are vital for effective data use, "combating a siege mentality and getting beyond blame and defensiveness to action are problems that go far beyond data use" (p18).

Herman and Gribbons' term 'efficacy' implies that teachers need to believe in change and their own capability and this theme recurs in the literature. Caputo and Rastelli's (2014) found that schools with high gains in student achievement were those which perceived themselves as effective in improving education. A belief among all staff that internal evaluation is beneficial and can lead to improvement has been found to be essential to its effectiveness (Bubb and Earley 2008; Lachat and Smith, 2005; Meuret and Morlaix, 2003; Schildkamp et al.,2012; Schildkamp and Viisscher,2013; Vanhoof and Van Petegem, 2011; Vanhoof et al., 2011). Gallimore et al. (2009) add that this is more likely when settings are stable with the flexibility needed from teachers emphasised by Honig and Ikemoto (2008).

Lack of belief in the benefits of internal evaluation is expressed in scepticism and resistance from teachers (Supovitz and Klein, 2003). Hall and Noyes (2007) describe a group of schools they term 'resisting' where internal evaluation is seen as a bureaucratic exercise where curriculum leaders are cynical about the process and other teachers complain about the workload and change fatigue. In these schools, internal evaluation is seen as a process which brings inspection and test- based accountability into the school, encouraging conformity and teaching to the test.

3.3.6 Accountability

Along with the need for extensive support for internal evaluation a degree of accountability has been found to be important in stimulating and sustaining engagement in the process. Simons (2013) reports that voluntary internal evaluation only works in committed schools and that prescription is required to get it underway, citing the example of Norway, where few schools undertook the process while it was optional. Janssens and Van Amelsvoort (2008) based on research in Denmark, England, Northern Ireland and Scotland, found that internal evaluation has a stronger position where there is an inspection framework to provide standards and expectations. Supovitz and Weathers (2004) found that not only did the peer review system they describe provide support for constructive discussion about improvements but it also created a sense of accountability and urgency.

In research conducted across England, Germany, Lithuania, the Netherlands and Poland, Schildkamp et al. (2014) found pressures from the accountability systems in use in each country influenced the use of data for decision making. Although both external and internal accountability incentives may be needed to use data to drive improvement (Honig and Ikemoto, 2008; Rallis and McMullen, 2000; Vanhoof et al., 2014; Schildkamp et al., 2013) it would seem that there must be a balance between external accountability and challenge and support for capacity building to inform transformational change MacBeath, 2004). Ozga et al. (2011) comment on tensions between what they term as 'hard' governance, composed of externally determined regulation, benchmarks and targets and 'soft' governance whereby a cycle of self- evaluation and improvement actions leads to a continuous process of educational improvement. Many of the negative, and unintended, effects of internal evaluation are reported to be associated with accountability beliefs. In research on the effects of school inspections in the Netherlands, Ehren et al (2014) found that expectations and educational standards set by an inspection framework help to drive internal evaluation and school improvement, but they also found, in a minority of schools, that there was a narrowing of the curriculum and "window dressing" of reports sent to the inspectorate to present a favourable picture of the school. In England, internal evaluation may be seen as an internal version of inspection as in some of the schools investigated by Hall and Noyes (2007). In these schools senior leaders made unannounced classroom observations and

judgements against inspection criteria, resulting in stress for curriculum leaders who found it hard to combine a monitoring role with supportive relationships and where teachers who expressed different views might be viewed as out of date or incompetent. Scepticism about the value of internal evaluation compared with the judgements made by external evaluators or through test- based accountability systems emerged in studies by, in the USA, (Jimerson, 2014; Supovitz and Weathers, 2004), Chile, (Montecinos et al., 2014) and Hong Kong (Yeung, 2011).

3.4 Mechanisms of effective internal evaluations

In emphasising necessary conditions, research implies that the mechanisms for effective internal evaluation are to be found in the activities of collecting and interpreting data, in the work of leaders and in the activities of collaborative, supportive groups of teachers who believe in and are committed to, the possibility of improvement for their school and the education of their students. There is comparatively little detail in the research literature about the exact nature of these activities, both in initiating and acting on internal evaluation and in embedding and institutionalising evaluation and improvement as a way of working for the school. Mechanisms that have been identified in the literature are discussed under the headings: accepting and interpreting feedback from internal evaluations, building capacity and organisational learning, implementing improvements.

3.4.1 Accepting and interpreting feedback.

It is perhaps because feedback is seen as a dialogue, in which information is sent from one party to another that the term occurs rarely in the literature on internal evaluation and only in relation to the feedback from an external tool for collecting and analysing data. Where data collection and analysis is a shared activity, the term is less appropriate.

Gaertner (2014) describes a system in Germany that collects feedback from students and feeds this back to individual teachers, with a key finding that it was only in those schools that discussed the feedback in collaborative groups that made organisational improvements. In another example where internal evaluation was supported through the use of an external tool in the Netherlands, Schildkamp et al. (2012) found that those schools where improvements were made had studied the feedback provided, discussed it and then taken measures to improve the quality of education. Less success was experienced through supported use of a tool designed in Flanders to collect and feedback on pupil performance and benchmarking data, where the use of this was limited, both in those schools that had had support and those that had not (Vanhoof et al., 2012).

3.4.2 Building capacity and organisational learning

The importance of support for schools for internal evaluation, both through the leadership of the principal and other school leaders, and external support of trainers, facilitators and critical friends, is emphasised throughout the literature. If it is to contribute to improvements in the educational experience of students in the longer term, then capacity must be built within the school and processes become institutionalised as a way of working, so that schools become learning organisations (Grek and Ozga,2012; Plowright, 2007) with ongoing collaborative discussion and decision making (Marsh and Farrell, 2015; Simons, 2013).

Dunn et al. (2013) describe a state- wide approach in which the need to build capacity was recognised from the start. The authors describe how a data system was established in Ontario, Canada. In the district taken as a case, there were already collaborative networks for school improvement. The system was introduced through a series of workshops which were designed to show participants how they could use their learning from data to engage in conversations with the rest of the staff in their schools. When they returned to their schools, participants worked with colleagues in professional learning communities to examine data about their existing practices. Additional support from a trained facilitator was added to the programme in the second year. The training and modelling of a process for collaborative enquiry in a professional learning community became a way that was collectively understood and shared.

Principals were responsible for promoting the process of collaborative learning from data and for monitoring changes in practice. The effectiveness of the approach is demonstrated across the 151 elementary and 32 secondary schools in the district where student achievement rose, and continued to rise, faster than elsewhere in the state. What appears to have been successful here is training and modelling of a protocol for working on data in professional learning communities. Ancess et al. (2007) and Cosner (2011) refer to protocols as 'structure' and 'tools and processes' that were developed to support focussed collaborative discussion about data. Gallimore et al. (2009) provide detail of an inquiry- based protocol used with grade- level teams in schools that were the subject of the research and which substantially improved student achievement. The protocol provided guidance on establishing a goal; planning; implementing; monitoring via common assessments; evaluating and moving to the next cycle. Similarly, Schildkamp and Ehren (2013) describe the structured approach used by data teams in ongoing research in the Netherlands, in which the steps are those of defining a problem; coming up with hypotheses about possible causes of the problem; collecting, analysing and interpreting relevant data; drawing conclusions and implementing plans to improve.

In New Zealand, internal evaluation, termed self review, is expected and monitored by the Educational Review Office, which also conducts inspections. For primary schools, there are no external tests, but schools are expected to assess students against national standards. In the first national evaluation of schools' evaluative capacity, Lai and Hsiao (2014) found that only about two – thirds of the school clusters investigated could produce high quality data for self review even when they had support. This implies that, despite the long- standing expectation in New Zealand for self review, capacity is not yet sufficient for using it as s process for sustainable and continuous school improvement. Lai and McNaughton (2013) and Timperley and Parr (2009) give separate examples of a process which aimed to improve the quality of self review in targeted schools that were part of a literacy intervention. As in the work described by Dunn et al., capacity building was fostered through the carefully planned steps of the intervention. In the first phase, with teachers working in clusters of schools, there was an initial discussion and analysis of data and the authors emphasise that it was important that this included, not only student achievement data, but also data gathered through lesson observations, student work and student surveys. Following the initial discussion, which was facilitated by a researcher, professional development was provided to meet the learning and teaching needs identified through discussion. Timperley and Parr add that this included professional development for leaders as well as for teachers. Professional content knowledge was frequently found to be a factor that needed to be addressed here. Professional learning communities within the clusters reviewed changes made and shared practice and it was through these that the learning was integrated into the normal routines of the school. This structured and supported approach to capacity building was successful as progress of students in Lai and McNaughton's study was greater than expected after three years, including for students that had joined the school subsequent to the project being first introduced.

In a further example, Farley- Ripple and Buttram (2014) describe a state- wide initiative to enhance internal evaluation and drive improvement in Delaware, USA. At state level, teachers were given a time allocation of 90 minutes per week during which they were expected to engage in professional learning communities and discuss and use data for improvement. The findings were that in only some of the schools had the opportunity for collaborative working been used to make instructional improvements. These schools, all in one district, differed from others in the state in the extent to which they conformed to a similar model to those reported as being successful by Dunn et al. and by Lai and McNaughton. There was a shared vision of the collaborative evaluation and improvement process, combined with support from the district, the active engagement of school leaders, norms and expectations for sharing information and working collaboratively and monitoring of effectiveness. In Honig and Ikemoto's (2008) report on internal evaluation in three school districts in the USA there was also just one of these that had had a shared vision for school improvement and provided professional development and tools for schools to use in collective identification of needs and goals and it was in this district that there was the greatest improvement in teaching and learning.

The example above suggest that, at state or district level, policy goals and actions can be used to build capacity for successful internal evaluation and the leadership role at a level above and individual school may be important (Wohlstetter et al., 2008) . On a smaller scale, Halverson et al. (2007) describe the routines in four elementary and middle schools in the USA that led to these schools being identified as having a strong track record of improvement in student achievement and of using data to guide decision making. In each school there was a belief in the value of using data to improve learning. Understandings of data were broad, with test data, student demographics, classroom observation, student work and student survey data all collected. Data was discussed in groups and "the discussions provided an occasion to develop shared understanding of purpose and strong professional community among the leadership team" (p 21). Opportunities for discussion were deliberately planned for at organisational level, so that staff met in teams to reflect collaboratively on data and use it to plan improvement at school level. In the more successful university – school partnership described by Herman and Gribbons (2001), similar processes for data use introduced with the help of the researcher became integrated into the school's way of working. The strength of routines and protocols for capacity building in internal evaluation is that they act as a store about the types of data that teachers need to notice and how they ought to discuss and interpret data, discuss implications for practice and plan for improvement (Kallemeyn, 2014). A different approach to building capacity is through a collaborative evaluation, such as that studied by Christie et al. (2004) through a case study of the evaluation of an externally-funded programme in a community college in California. An evaluation team composed of both external and faculty members developed through shared meetings into a learning community, such that "Evaluation activities have become an essential contributor to the college's understanding of learning outcomes and what transpires in the classroom" (p 130). Members of the team have developed their evaluation expertise, so that by the fourth year of the research, the external evaluator had a lower profile in the evaluation team, with responsibility for leading activity handed over to faculty members.

Another approach to capacity building is through the collaborative evaluation described by Christie et al. (2004) of an externally- funded project in a community college in California, USA. An evaluation team was formed to include the external evaluator and members of the college faculty and this worked together to develop an evaluation approach tailored to the needs of the college and the external funders. The team developed as a learning community and the quality of their evaluation changed attitudes in the college as its impact on learning was seen and became more valued. Evaluation capacity grew, with the external evaluator's role diminishing over time.

Perhaps surprisingly, considering the investment made by partners and policy makers in introducing and establishing internal evaluation in schools which is a feature of the studies discussed above, only one example has been found which mentions succession planning. In the Welsh 'high reliability project' schools visited by Stringfield et al. (2008), five years after the project had first been introduced, the majority of schools continued to use the project principles and continued making strong academic progress. Staff in schools were continuously re-engaged through shared residential workshops and professional development designed to share good practice and barriers and to discuss implications for future improvement. Visits within and between schools were used so that teachers could observe and discuss each other's lessons. Leadership succession planning was in place to ensure that incoming school principals were familiar with project principles.

Protocols for internal review that are reinforced by working with other schools using the same processes, at small group, district or state level are most frequently stimulated by external support. In addition, the responsibility of the school principal is mentioned by several authors to build capacity for internal evaluation. In systems where there are structures and traditions of internal school groups, these groups can support discussion and use of data for improvement (Schildkamp et al., 2014). Where these are not historically in place and where there is no external driver, individual school leaders may need to build collaboration, including by providing a time allocation. (Dembosky et al., 2006; Lachat and Smith, 2005). Principals may also need to be actively engaged in developing internal protocols and artefacts, such as data reports, to help keep school collaborative groups on track in discussion (Cosner, 2011; Earl, 2008; Young, 2006).

3.4.3 Implementing improvements

If internal evaluation is to lead to improvement, then plans developed as a result of the evaluation process must be implemented and monitored. Strong leadership to ensure that this takes place has been referred to above in the section on conditions for successful internal evaluation. McNaughton et al (2012) point out that data may be used to develop context specific teaching but that this needs to be put into effect to result in achievement gains for students. Anderson et al. (2010) found that it is the appropriateness of actions actually taken based on data-informed decisions that has an impact on the quality of teaching and learning.

Devos and Verhoeven (2003) provide an interesting example from Flanders where a process of internal evaluation of organisational climate was conducted through a partnership of each school with external researchers. In each of the case study schools, the schools agreed in discussion with the researchers about the implications of the results for their schools and on the changes that might be made to improve organisational climate. However, change occurred in none of the case study schools, due in two cases to the authoritarian leaders of the schools, who were either unable or unwilling to change. In the third school, it had been decided collaboratively that the proposed change would detract from the work of the school. The team conclude that the external perspective provided by the researchers was helpful in identifying what would otherwise have been blindspots for self- evaluators, but that change is dependent on internal factors, particularly on the leadership of the school.

4. Conclusion and discussion

The purpose of this review was to identify and summarise findings from international empirical and non- empirical research on the impact of internal evaluation in schools. In reviewing literature across different countries and cultures, what is most remarkable is the consistency among researchers and commentators about the conditions, mechanisms and challenges associated with successful internal evaluation. As data systems have become more sophisticated and benchmarking data and electronic feedback tools have become more widely available, the terminology of internal evaluation has changed, with "data use", "data teams" becoming common outside, as well as inside, the USA to cover activity in which information is collected, analysed and interpreted to inform the solution of an educational problem defined by the school. "Inquiry" or "research- informed" are other terms, used particularly when internal evaluation becomes a regular event, with successive evaluations building on the one before in a "cycle of inquiry". However, to return to Simons (2013) definition, the activity of internal evaluation as *a process of purposive evaluation of school practices which provides insights into the educational experiences of students, as more than those measured by test data* is consistent, whatever terminology is used.

4.1 The effects of internal evaluation.

The review has found that internal evaluation can have a positive effect on students' learning experiences and on their academic achievement, with evidence of improvement demonstrated in improved test scores and validation by external evaluation. Internal evaluation, in itself, may not lead directly to improvement actions but it does so through increasing the sensitivity of teachers to the conditions for learning in their school, which in turn leads to planning and action for school improvement and for better conditions for learning. Internal evaluation encourages school to use a range of methods of data collection, such as lesson observation, analysis of student work and collecting feedback from students and parents. Discussion of findings helps to identify professional development needs and informs goals and actions for school improvement planning.

Literature on internal evaluation is, on the whole, positive about its effects. The negative effects of internal evaluation referred to in the literature are, in the main, related to two factors:
- the pressure of external accountability and external quality assurance processes
- the quality of leadership in the school.

Pressures of external accountability may restrict understandings of student achievement, such that broader aims and goals of education are neglected in the collections of data (Simons, 2013) and lead to cynicism about the process and its value compared with external inspection. Unduly rigid frameworks for internal evaluation can lead to work overload and stress as well as cynicism and resistance. However, the stimulus of being held accountable for their ability to self-evaluate appears to be motivating for schools to engage in the process and to implement changes.

Leadership of the school, particularly from the school principal, is essential for mediating external accountability demands and promoting a culture in which internal evaluation is seen as an opportunity to learn, which can be shaped to fit shared values and aims of the organisation. Leaders, both in the school and at district, board or local authority level, can also do much to provide the resources including access to external support and time which are required for internal evaluation to be successful.

4.2 The conditions for successful internal evaluation

Internal evaluation is a complex and difficult activity, which requires schools to engage in questioning their existing practices and to reflect on the underpinning values that inform teaching and learning. It is not unexpected, then, that almost all of the studies identified for this review identified a number of essential conditions for successful internal evaluation. What is more surprising is the commonality and consistency in the conditions noted by researchers in jurisdictions with completely different accountability frameworks and cultures. These conditions appear to be necessary, whether or not internal accountability is mandated or optional, and whether or not there is external evaluation of schools. Essential conditions identified are:

- *Evaluation literacy*. Schools must know how to identify goals and questions for the internal evaluation, they must be aware of what data might be useful and of how this may be accessed, they must know how to analyse and interpret the data they have and, if it is to lead to improvement, they must be able to use their interpretation to plan for changes to existing practices.

- *Resources*. Time is the most frequently- mentioned resource need, in the short- term, for teachers to meet to plan collaboratively for the evaluation, to collect, analyse and interpret data and to plan and implement improvements. It is also needed in the longer term, so that expertise in evaluation literacy and evaluation processes may be spread throughout the school and become embedded into schools' practices. Systems for managing data and the availability of training in its use are also considered essential. Self- evaluation tools, the availability of benchmarking data and data technologies were mentioned by some authors.

- *Leadership and a supportive culture*. The most significant leadership role for successful internal evaluation is that of the principal, with distribution of leadership and shared ownership vital as processes of internal evaluation develop. Leadership at district, school board or local authority level is also a feature of many of the examples where effective practice in internal evaluation has become widespread and sustainable. Leadership is essential for providing and sustaining a vision and promoting a *supportive culture*, where there is trust and shared belief that education can be improved by the teachers in the school. This is particularly important where external quality assurance methods, including inspection, are threatening and stressful for teachers. Leadership is essential for ensuring that necessary resources are in place.

- *External support and accountability.* Although external accountability may be threatening and lead to negative effects, the literature suggests that the need to be accountable, whether to a national or regional mandate or to the expectations of partners and stakeholders, is a necessary stimulus to engage in internal evaluation. External support for developing evaluation literacy is mentioned most frequently, through training but also through facilitation of discussion and interpretation of data, so that it may lead

to planning for improvement. External viewpoints help to provide objectivity to validate the findings of internal evaluation and to provide challenge.

The most successful examples of internal evaluation are those which have been attentive to the need to *build capacity and organisational learning*. It is through embedding internal evaluation into the way of working of the school that learning and improvement can be secured and sustained. Capacity building seems to be most effective when skilled experts work alongside a school, or group of schools, over a period of time, in such a way that vision and goals can be shared and support put in place which is tailored to the needs of individual schools or school groupings. Facilitators working with groups of teachers, to model the use of data and questioning and to keep teachers focused on analysis and interpretation, seem to be particularly helpful. A *structured approach*, through the following of ordered steps, helps groups work methodically and provides a stored routine to contribute to organisational learning and the development of the school as a learning organisation. Structured approaches, such as protocols for discussion, are also helpful in guiding conversations about student work or other data, so that they are both critical and challenging. Agreed supportive tools, such as data reports and data collection tools, such as learning walks are further aids to capacity building and organisational learning. Although only mentioned in one study (Stringfield et al., 2008) succession planning, particularly for school principals would seem to be an essential mechanism to secure sustained capacity.

To conclude, it must be noted that internal evaluation takes place in schools which each operate within a broader and unique culture. The purposes and aims of internal evaluation in each school requires that a balance must be struck between the vision and goals of the school and those of the culture in which it operates. Exactly where and how will depend on the context of the school, the culture in which it operates and the values and views of its leader, its students and of other stakeholders. Or, as Macbeath (2004) expresses it, "The criteria we use to evaluate learning, teaching, ethos and leadership form a delicate mosaic, reflecting different interests, needs and imperatives." (p 8).

Ozga et al. (2011) refer to tensions between what they term as 'hard' governance, composed of externally determined regulation, benchmarks and targets and 'soft' governance whereby a cycle of self- evaluation and improvement actions leads to a continuous process of educational improvement. Although they point to these tensions as reflecting a universal trend towards uniformity and conformity in education, they acknowledge that there are differences between jurisdictions dependent on their culture and history of educational reform. However, the pressures of external regulation may be particularly demanding for schools with a history of underperformance even when compared with other, more advantaged schools in the same jurisdiction and the same overarching culture (Diamond and Cooper, 2007). Although the evidence about necessary conditions and mechanisms for successful evaluation appears to be both universal and convincing, the success of internal evaluation is dependent on the context and circumstances in which the school does its work.

Bibliography

Ah-Teck, J. C., & Starr, K. C. (2014). Total Quality Management in Mauritian education and principals' decision-making for school improvement. *Journal of Educational Administration, 52*(6), 833-849.

Ancess, J., Barnett, E., & Allen, D. (2007). Using Research to Inform the Practice of Teachers, Schools, and School Reform Organizations. *Theory Into Practice, 46*(4), 325-333.

Andersen, V. N., Dahler-Larsen, P., & Pedersen, C. S. m. (2009). Quality assurance and evaluation in Denmark. *Journal of Education Policy, 24*(2), 135-147.

Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading Data Use in Schools: Organizational Conditions and Practices at the School and District Levels. *Leadership and Policy in Schools, 9*(3), 292-327.

Barrett, T. (2009). *Teacher Conversation and Community: What Matters in Smaller Learning Communities and Inquiry-Based High School Reform.* (Ph. D.), New York University.

Blok, H., Sleegers, P., & Karsten, S. (2008). Looking for a Balance between Internal and External Evaluation of School Quality: Evaluation of the SVI Model. *Journal of Education Policy, 23*(4), 379-395.

Bubb, S., & Earley, P. (2008). *From self-evaluation to school improvement: the importance of effective staff development*: CfBT Education Trust.

Caputo, A., & Rastelli, V. (2014). School improvement plans and student achievement: Preliminary evidence from the Quality and Merit Project in Italy. *Improving Schools, 17*(1), 72-98.

Chapman, C. (2000). Improvement, Inspection and Self-Review. *Improving Schools, 3*(2), 57-63.

Christie, C. A., Ross, R. M., & Klein, B. M. (2004). Moving toward collaboration by creating a participatory internal-external evaluation team: A case study. *Studies in Educational Evaluation, 30*(2), 125-134.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of Evidence Use in School Districts: Mapping the Terrain. *American Journal of Education, 112*(4), 469=495.

Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis, 29*(4), 469-483.

Cosner, S. (2011). Teacher Learning, Instructional Considerations and Principal Communication: Lessons from a Longitudinal Study of Collaborative Data Use by Teachers. *Educational Management Administration & Leadership, 39*(5), 568-589.

Cowan, G. (2008). *New relationship with schools evaluation*: Department for Children, Schools and Families.

Croxford, L., Grek, S., & Shaik, F. J. (2009). Quality Assurance and Evaluation (QAE) in Scotland: Promoting Self-Evaluation within and beyond the Country. *Journal of Education Policy, 24*(2), 179-193.

Datnow, A. C., Marisa. (2000). Teachers' Responses to Success for All: How Beliefs, Experiences, and Adaptations Shape Implementation. *American Educational Research Journal, 37*(3), 775-799.

Davidsdottir, S., & Lisi, P. (2007). Effects of Deliberative Democracy on School Self-Evaluation. *Evaluation, 13*(3), 371-386.

Dembosky, J. W., Pane, J. S., Barney, H., & Christina, R. (2006). Data driven decision making in Southwestern Pennsylvania school districts: RAND Corporation.

Demetriou, D., & Kyriakides, L. (2012). The impact of school self-evaluation upon student achievement: a group randomisation study. *Oxford Review of Education, 38*(2), 149-170.

Demie, F. (2003). Using Value-Added Data for School Self-Evaluation: A Case Study of Practice in Inner-City Schools. *School Leadership & Management, 23*(4), 445-467.

Devos, G., & Verhoeven, J. C. (2003). School Self-Evaluation-Conditions and Caveats. The Case of Secondary Schools. *Educational Management & Administration, 31*(4), 403-420.

Diamond, J. B., & Cooper, K. (2007). chapter 10 The Uses of Testing Data in Urban Elementary Schools: Some Lessons from Chicago. *Yearbook of the National Society for the Study of Education 106.1 (2007): 241-263., 106*(1), 241-263.

Earl, L. (2008). Leadership for Evidence- Informed Conversations. In L. Earl & H. Timperley (Eds.), *Professional Learning Conversations: Challenges in Using Evidence for Improvment*. The Netherlands: Springer.

Earley, P., & Bubb, S. (2014). Data and Inquiry Driving School Improvement: Recent Developments

in England. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal), 1*(9), 167-184.

Education, S. (2012). Quality and Improvement in Scottish Education: Trends in Inspection Findings 2008-2011: Education Scotland.

Ehren, C. M., & Shackleton, N. (2014). Impact of school inspections on teaching and learning in primary and secondary education in the Netherlands. London: Institute of Education.

Ehren, M., Perryman, J., & Shackleton, N. (2014). Setting expectations for good education: how Dutch school inspections drive improvement. *School Effectiveness and School Improvement*, 1-32.

Emstad, A. B. (2011). The Principal's Role in the Post-Evaluation Process.--How Does the Principal Engage in the Work Carried out after the Schools Self-Evaluation? *Educational Assessment, Evaluation and Accountability, 23*(4), 271-288.

Estyn (2012). *How well are the All-Wales Core Data Sets used to inform self-evaluation and planning for improvement?* Cardiff: Her Majesty's Inspectorate for Education and Training in Wales (Estyn).

Farley-Ripple, E. N., & Buttram, J. L. (2014). Developing collaborative data use through professional learning communities: Early lessons from Delaware. *Studies in Educational Evaluation, 42*, 41-53.

Farrell, C. C. (2014). Designing School Systems to Encourage Data Use and Instructional Improvement: A Comparison of School Districts and Charter Management Organizations. *Educational Administration Quarterly*, 1-34.

Gaertner, H. (2014). Effects of student feedback as a method of self-evaluating the quality of teaching. *Studies in Educational Evaluation, 42*, 91-99.

Gallimore, R., Ermeling, R. M., Saunders, W. M., & Goldenberg, C. N. (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. *Elementary School Journal, 537-553*(5), 133- 157.

Geijsel, F. P., Kruger, M. L., & Sleegers, P. J. (2010). Data Feedback for School Improvement: The Role of Researchers and School Leaders. *Australian Educational Researcher, 37*(2), 59-75.

Gray, J., Croxford, L., Pederson, C. S., Rinne, R., Silmari- Salo, S., Simola, H., & Makinen- Streng, M. (2011). Teachers' perceptions of quality assurance and evaluation. In J. Ozga, P. Dahler-Larsen, H. Simola, & C. Segerholm (Eds.), *Fabricating Quality in Education: Data and Governance in Europe*. London: Routledge.

Grek, S., & Ozga, J. (2011). Governing learning: self-evaluation of schools in Scotland. *Education et sociétés, 28*, 77-91.

Hall, C., & Noyes, A. (2007). *The impact of school self evaluation processes on British teachers' views of their work and professionalism*. Paper presented at the BERA, London.

Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The New Instructional Leadership: Creating Data-Driven Instructional Systems in School. *Journal of School Leadership, 17*(2), 159-194.

Hartong, S. (2012). Overcoming resistance to change: PISA, school reform in Germany and the example of Lower Saxony. *Journal of Education Policy, 27*(6), 747-760.

Herman, J. L., & Gibbons, B. (2001). Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation. . Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California.

Hofman, R. H., de Boom, J., & Hofman, W. H. A. (2010). Quality control in primary schools: progress from 2001–2006. *School Leadership & Management, 30*(4), 335-350.

Hofman, R. H., Dijkstra, N. J., & Hofman, H. W. (2008). Internal versus External Quality Management. *International Journal of Leadership in Education, 11*(3), 281-300.

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.

Honig, M. I., & Ikemoto, G. S. (2008). Adaptive Assistance for Learning Improvement Efforts: The Case of the Institute for Learning. *Peabody Journal of Education, 83*(3), 328-363.

Ikemoto, G., & Marsh, J. A. (2007). Cutting through the "Data-driven" Mantra: Different Conceptions of Data-Driven Decision Making: RAND Corporation.

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.

Jaquith, A., & McLaughlin, M. (2010). A Temporary, Intermediary Organization at the Helm of

Regional Education Reform: Lessons from the Bay Area School Reform Collaborative. 85-103.

Jimerson, J. B. (2014). Thinking about data: Exploring the development of mental models for "data use" among teachers and school leaders. *Studies in Educational Evaluation, 42*, 5-14.

Kallemeyn, L. M. (2014). School-level organizational routines for learning: supporting data use. *Journal of Educational Administration, 52*(4), 529-548.

Karagiorgi, Y. (2012). School self-evaluation at an embryonic stage: depicting teachers' experiences with a participative project. *International Journal of Leadership in Education: theory and practice, 15*(2), 199-220.

Karagiorgi, Y., Nicolaidou, M., Yiasemis, C., & Georghiades, P. (2015). Emergent data-driven approaches to school improvement: The journey of three schools through self-evaluation. *Improving Schools, 18*(1), 69-82.

Kerr, K., Marsh, J., Ikemoto, G., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. . *American Journal of Education, 112*(4), 496-520.

Knapp, M. S., & Feldman, S. B. (2012). Managing the Intersection of Internal and External Accountability: Challenge for Urban School Leadership in the United States. *Journal of Educational Administration, 50*(5), 666-694.

Kyriakides, L., & Campbell, R. J. (2004). School self-evaluation and school improvement: A critique of values and procedures. *Studies in Educational Evaluation, 30*(1), 23-36.

Lachat, M. A., & Smith, S. (2005). Practices That Support Data Use in Urban High Schools. *Journal of Education for Students Placed at Risk (JESPAR), 10*(3), 333-349.

Lai, M., & Kushner, S. (2013). *A Developmental and Negotiated Approach to School Self-evaluation.* (Vol. 14): Emerald Group Publishing Limited.

Lai, M. K., & Hsiao, S. (2014). Developing data collection and management systems for decision-making: What professional development is required? *Studies in Educational Evaluation, 42*, 63-70.

Lai, M. K., & McNaughton, S. (2013). Analysis and Discussion of Classroom and Achievement Data to Raise Student Achievement In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based Decision Making in Education: Challenges and Opportunities*. Dordrecht: Springer Science+Business Media

Lai, M. K., McNaughton, S., Timperley, H., & Hsiao, S. (2009). Sustaining continued acceleration in reading comprehension achievement following an intervention. *Educational Assessment, Evaluation and Accountability, 21*(1), 81-100.

Lasky, S., Schaffer, E. C., & Hopkins, T. (2008). Learning to Think and Talk from Evidence: Developing System-wide Capacity for Learning Conversations. In L. Earl & H. Timperley (Eds.), *Professional Learning Conversations: Challenges in Using Evidence for Improvement*. The Netherlands: Springer.

Leung, C. K. L. (2005). *Accountability Versus School Development: Self-Evaluation in an International School in Hong Kong.* (M Ed M Ed), The University of Hong Kong, Hong Kong. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=20h&AN=16757471&site=ehost-live Available from EBSCOhost 20h database.

Levin, J. A., & Datnow, A. (2012). The principal role in data-driven decision making: using case-study data to develop multi-mediator models of educational reform. *School Effectiveness and School Improvement, 23*(2), 179-201.

Little, J. W. (2012). Understanding Data Use Practice among Teachers: The Contribution of Micro-Process Studies. *American Journal of Education, 118*(2), 143-166.

Little, J. W., & Curry, M. (2008). Structuring Talk about Teaching and Learning: The Use of Evidence in Protocol-Based Conversation. In L. Earl & H. Timperley (Eds.), *Professional Learning Conversations: Challenges in Using Evidence for Improvement*. The Netherlands: Springer.

MacBeath, J. (2004). Putting the Self Back into Self-Evaluation. *Improving Schools, 7*(1), 87-91.

Marsh, J., Hamilton, L., & Gill, B. (2008). Assistance and Accountability in Externally Managed Schools: The Case of Edison Schools, Inc. *Peabody Journal of Education, 83*(3), 423-458.

Marsh, J., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy, 24*(6), 872–907.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. . *Teachers College Record, 114*(11), 1- 48.

Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership, 43*(2), 269-289.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). Making Sense of Data-Driven Decision Making in Education: Rand Education.

McDougall, D., Saunders, W. M., & Goldenberg, C. (2007). Inside the Black Box of School Reform: Explaining the how and why of change at Getting Results schools. *International Journal of Disability, Development and Education, 54*(1), 51-89.

McNamara, G., & O'Hara, J. (2005). Internal Review and Self-Evaluation--The Chosen Route to School Improvement in Ireland? *Studies in Educational Evaluation, 31*(4), 267-282.

McNamara, G., & O'Hara, J. (2006). Workable compromise or pointless exercise? School-based evaluation in the Irish context. *Educational Management Administration and Leadership, 34*(4), 564-582.

McNamara, G., & O'Hara, J. (2008). The Importance of the Concept of Self-Evaluation in the Changing Landscape of Education Policy. *Studies in Educational Evaluation, 34*(3), 173-179.

McNamara, G., & O'Hara, J. (2012). From looking at our schools (LAOS) to whole school evaluation -management, leadership and learning (WSE-MLL): the evolution of inspection in irish schools over the past decade. *Educational Assessment, Evaluation & Accountability, 24*(2), 79-97.

McNamara, G., O'Hara, J., Lisi, P. L., & Davidsdottir, S. (2011). Operationalising self-evaluation in schools: experiences from Ireland and Iceland. *Irish Educational Studies, 30*(1), 63-82.

McNaughton, S., Lai, M. K., & Hsiao, S. (2012). Testing the effectiveness of an intervention model based on data use: a replication series across clusters of schools. *School Effectiveness and School Improvement, 23*(2), 203-228.

Means, B., Padilla, C., & Gallagher, L. (2010). Use of education data at the local level: From accountability to instructional improvement. Washington DC: Department of Education, Office of Planning, Evaluation, and Policy Development.

Meuret, D., & Morlaix, S. (2003). Conditions of Success of a School's Self-Evaluation: Some Lessons of an European Experience. *School Effectiveness and School Improvement, 14*(1), 53-71.

Montecinos, C., Madrid, R., Fernández, M., & Ahumada, L. (2014). A goal orientation analysis of teachers' motivations to participate in the school self-assessment processes of a quality assurance system in Chile. *Educational Assessment, Evaluation & Accountability, 26*(3), 241-261.

Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using Student-Assessment Results to Improve Instruction: Lessons From a Workshop. *Journal of Education for Students Placed at Risk (JESPAR), 10*(3), 269-280.

Mutch, C. (2012). Complementary Evaluation: The Development of a Conceptual Framework to Integrate External and Internal Evaluation in the New Zealand School Context. *Policy Futures in Education, 10*(5), 569-586.

Neil, P., & Johnston, J. (2005). An approach to analysing professional discourse in a school self-evaluation project. *Research in Education, 73*(-1), 73-86.

Nevo, D. (2001). School evaluation: internal or external? *Studies in Educational Evaluation, 27*, 95-106.

O'Brien, S., McNamara, G., & O'Hara, J. (2014). Critical facilitators: External supports for self-evaluation and improvement in schools. *Studies in Educational Evaluation*.

O'Brien, S., McNamara, G., & O'Hara, J. (2015). Supporting the consistent implementation of self-evaluation in Irish post-primary schools. *Educational Assessment, Evaluation & Accountability, not yet assigned*.

Ozga, J. (2009). Governing education through data in England: from regulation to self-evaluation. *Journal of Education Policy, 24*(2), 149-162.

Ozga, J. (2011). *Fabricating quality in education : data and governance in Europe*. London: Routledge.

Park, V., & Datnow, A. (2008). Collaborative Assistance in a Highly Prescribed School Reform Model: The Case of Success for All. *Peabody Journal of Education, 83*(3), 400-422.

Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: district and school connections in data-driven decision-making. *School Leadership & Management, 29*(5), 477-494.

Pedersen, C. S. m. (2008). *QAE and the Teaching Profession in Denmark*. Paper presented at the European Conference on Educational Research (ECER) Network 23 Goteborg.

Plowright, D. (2007). Self-evaluation and Ofsted Inspection. Developing an Integrative Model of School Improvement. *Educational Management Administration & Leadership, 35*(3), 373-393.

Porter, K. E., & Snipes, S. C. (2006). The challenge of supporting change: Elementary student achievement and the Bay Area School Reform Collaborative's focal strategy. New York: MDRC.

Quint, J., Sepanik, S., & Smith, J. (2008). Using student data to improve teaching and learning: Findings from an evaluation of the Formative Assessments of Student Thinking in Reading (FAST-R) program in Boston elementary schools. New York: MDRC.

Rallis, S. F., & MacMullen, M. M. (2000). Inquiry-minded schools: Opening doors for accountability. . *Phi Delta Kappan, 81*(10), 766-773.

Ritchie, R. (2002). School improvement in the context of a primary school in special measures. *Teacher Development, 6*(3), 329-346.

Ryan, K. E., Chandler, M., & Samuels, M. (2007). WHAT SHOULD SCHOOL-BASED EVALUATION LOOK LIKE? *Studies in Educational Evaluation, 33*(3–4), 197-212.

Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Increasing Achievement by Focusing Grade-Level Teams on Improving Classroom Learning: A Prospective, Quasi-Experimental Study of Title I Schools. *American Educational Research Journal, 46*(4), 1006-1033.

Schildkamp, K., & Ehren, M. C. M. (2013). From "Intuition"- to "Data"-based Decision Making in Dutch Secondary Schools? In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based Decision Making in Education: Challenges and Opportunities*
Dordrecht: Springer Science+Business Media

Schildkamp, K., Karbautzki, L., & Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. *Studies in Educational Evaluation, 42*, 15-24.

Schildkamp, K., Vanhoof, J., Petegem, P. v., & Visscher, A. J. (2012). The use of school self-evaluation results in the Netherlands and Flanders. *British Educational Research Journal, 38*(1), 125-152.

Schildkamp, K., & Visscher, A. (2009). Factors influencing the utilisation of a school self-evaluation instrument. *Studies in Educational Evaluation, 35*(4), 150-159.

Schildkamp, K., Visscher, A., & Luyten, H. (2009). The effects of the use of a school self-evaluation instrument. *School Effectiveness and School Improvement, 20*(1), 69-88.

Simons, H. (2013). Enhancing the quality of education through school self- evaluation In M. K. Lai & S. Kushner (Eds.), *A Developmental and Negotiated Approach to School Self-evaluation. Advances in Program Evaluation Volume 14*. Bingley, UK: Emerald Group Publishing.

Sjobakken, O. J., & Dobson, S. (2013). School self- evaluation in the longer time scale. Experiences from a small Scandinavian state In M. Lai & S. Kushner (Eds.), *A Developmental and Negotiated Approach to School Self-evaluation. Advances in Program Evaluation Volume 14*. Bingley, UK: Emerald Group Publishing.

Stringfield, S., Reynolds, D., & Schaffer, E. C. (2008). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the High Reliability Schools project. *School Effectiveness and School Improvement, 19*(4), 409-428.

Supovitz, J., & Klein, V. (2003). Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement. Pennsylvania: Consortium for Policy Research in Education, University of Pennsylvania, Graduate School of Education.

Supovitz, J., & Weathers, J. M. (2004). Dashboard lights: Monitoring implementation of district instructional reform strategies. Philadelphia: Consortium for Policy Research in Education.

Sutherland, S. (2004). Creating a Culture of Data Use for Continuous Improvement: A Case Study of an Edison Project School. *The American Journal of Evaluation*.

Swaffield, S. (2004). Critical friends: supporting leadership, improving learning. *Improving Schools, 7*(3), 267-278.

Swaffield, S., & MacBeath, J. (2005). School self-evaluation and the role of a critical friend. *Cambridge Journal of Education, 35*(2), 239-252.

Timperley, H. (2013). The New Zealand educational context: evaluation and self- review in a self-managing system. In M. K. Lai & S. Kushner (Eds.), *A Developmental and Negotiated Approach to School Self-evaluation. Advances in Program Evaluation Volume 14*. Bingley, UK: Emerald Group Publishing.

Timperley, H. S., & Parr, J. M. (2009). Chain of influence from policy to practice in the New Zealand literacy strategy. *Research Papers in Education, 24*(2), 135-154.

Trachtman, R. (2007). Inquiry and Accountability in Professional Development Schools. *The Journal of Educational Research. 100*(4), 197-203.

Van Petegem, P., & Vanhoof, J. (2007). Towards a Model of Effective School Feedback: School Heads' Points of View. *Educational Research and Evaluation, 13*(4), 311-325.

Vanhoof, J., De Maeyer, S., & Van Petegem, P. (2011). Variation in the conduct and the quality of self-evaluations: a multi-level path analysis. *Educational Studies, 37*(3), 277-287.

Vanhoof, J., & Van Petegem, P. (2010). Evaluating the quality of self-evaluations: The (mis)match between internal and external meta-evaluation. *Studies in Educational Evaluation, 36*(1-2), 20-26.

Vanhoof, J., & Van Petegem, P. (2011). Designing and evaluating the process of school self-evaluations. *Improving Schools, 14*(2), 200-212.

Vanhoof, J., Vanlommel, K., Thijs, S., & Vanderlocht, H. (2014). Data use by Flemish school principals: impact of attitude, self-efficacy and external expectations. *Educational Studies, 40*(1), 48-62.

Vanhoof, J., Verhaeghe, G., Van Petegem, P., & Valcke, M. (2012). Flemish Primary Teachers' Use of School Performance Feedback and the Relationship with School Characteristics. *Educational Research, 54*(4), 431-449.

Vanhoof, J., Verhaeghe, G., Verhaeghe, J. P., Valcke, M., & Van Petegem, P. (2011). The Influence of Competences and Support on School Performance Feedback Use. *Educational Studies, 37*(2), 141-154.

Vázquez, M. I., & Gairín, J. (2013). Institutional self-evaluation and change in educational centres. *International Journal of Leadership in Education, 17*(3), 327-352.

Verhaeghe, G., Schildkamp, K., Luyten, H., & Valcke, M. (2015). Diversity in school performance feedback systems. *School Effectiveness and School Improvement*, 1-27.

Wayman, J. C., Cho, V., & Johnston, M. C. (2007). The Data-Informed District: A District-Wide Evaluation of Data Use in the Natrona County School District. Austin, Texas: University of Texas.

Wayman, J. C., & Stringfield, S. (2006). Technology‐Supported Involvement of Entire Faculties in Examination of Student Data for Instructional Improvement. *American Journal of Education, 112*(4), 549-571.

Webb, R., Vulliamy, G., Häkkinen, K., & Hämäläinen, S. (1998). External Inspection or School Self-evaluation? A Comparative Analysis of Policy and Practice in Primary Schools in England and Finland. *British Educational Research Journal, 24*(5), 539-556.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a System for Data-Driven Decision-Making: Applying the Principal-Agent Framework. *School Effectiveness and School Improvement, 19*(3), 239-259.

Wong, M. N., & Li, H. (2010). From External Inspection to Self-Evaluation: A Study of Quality Assurance in Hong Kong Kindergartens. *Early Education and Development, 21*(2), 205-233.

Wroe, A., & Halsall, R. (2001). School self-evaluation: Measurement and reflection in the school improvement process. *Research in Education, 65*(-1), 41-52.

Yeung, S. Y. S. (2011). A School Evaluation Policy with a Dual Character: Evaluating the School Evaluation Policy in Hong Kong from the Perspective of Curriculum Leaders. *Educational Management Administration & Leadership, 40*(1), 37-68.

Young, V. M. (2006). Teachers' Use of Data: Loose Coupling, Agenda Setting, and Team Norms. . *American Journal of Education, 112*(4), 521-548.

Annex. Evidence tables

Table 1. Studies reporting on effects and side effects of internal evaluations

	Type of effect					
	no effect	reflection on school quality and intentions to improve	school improvement	improved student achievement	improved conditions of learning	unintended effect
Country	Cyprus, Mauritius, New Zealand, Scotland, the Netherlands, Uruguay, USA, Wales	Denmark/ England/ Finland/Scotland/ Sweden (1 study), international (1 study), Cyprus, England (2 studies), Hong Kong (2 studies), Iceland, Ireland, Ireland/Iceland, Israel, Northern Ireland, The Netherlands, The Netherland/Belgium-Flanders	Cyprus, England (4 studies), Hong Kong, Italy, Northern Ireland, Scotland, The Netherlands, The Netherlands/Belgium-Flanders, USA (2 studies), Wales	Canada, Cyprus, England, Italy, New Zealand (3 studies), USA (3 studies), Wales	England (2 studies), England/Finland, Germany, Hong Kong, The Netherlands, USA (7 studies), New Zealand (2 studies)	Denmark, England (3 studies), Germany, Hong Kong (2 studies), Scotland, Uruguay, international
Year in which studies were reported	2008 (2 studies), 2009, 2012 (3 studies),2014 (2 studies)	2001 (2 studies), 2002, 2005, 2007, 2008, 2011 (3 studies), 2012, 2013, 2014, 2015 (2 studies)	2003 (2 studies), 2005 (2 studies), 2007, 2008 (3 studies), 2012 (3 studies), 2014 (2 studies), 2015	2008 (2 studies), 2009 (2 studies), 2010, 2011, 2012 (2 studies), 2013 (2 studies), 2014	1998, 2000, 2001, 2003, 2005, 2006 (2 studies), 2008, 2009 (2 studies), 2010 (2 studies), 2012 (2 studies), 2014	2001 (2 studies), 2007, 2008 (2 studies), 2009, 2010, 2012, 2013, 2014

Summaries	Ah Teck and Starr (2014) informal methods preferred are those likely to influence a low degree of change (Mauritius).	Davidsdottir and Lisi (2007) Researcher findings include changes in teacher perceptions of administration facilitating their professional growth; systematic data collection; teacher ownership and collaboration; and shared decision-making for improvement efforts in the schools (Iceland).	Bubb and Earley (2008) Completion of the SEF helped schools in focusing their improvement priorities (England)	Caputo and Rastelli (2014) High student achievement gains following training in self evaluation and good- quality school improvement plans, low gains in trained schools with poor-quality plans (Italy)	Chapman (2000) Peer observations, which were supportive and non-pressurized, led to improvements (England).	Andersen et al (2009) Measure fixation happens when practitioners focus on exactly what is being measured as an indicator of quality, often at the expense of genuine quality.. under advanced measure fixation, the indicator provides a definition of quality along with an indicator of how to measure quality. With advanced measure fixation it is not possible to demonstrate a cleavage between genuine quality and quality measured by an indicator, since the latter helps define the former (Denmark).
	Blok et al. (2008)completed SSEs were of low quality, often failing to answer questions set at the onset of the process,at the time of the research (2003-6) their completion was unlikely to lead to school improvement (The Netherlands)	Davies and Rudd (2001)The SSE helped the school to 'know where they were' prior to an Ofsted inspection.and helped them in linking their school improvement plans to local authority education plans. It can bring about a change in culture, including increased use of classroom observation; increased CPD; ownership of change; community	Caputo and Rastelli (2014) SIPs of schools with very high student achievement improvement provide a more accurate analysis of the context and a greater originality and specificity of improvement goals, whilst the SIPs of schools with very low student achievement improvement show a more precise and articulated diagnosis of student needs.' (p83) (pp 84,87) '..schools with very high student achievement improvement tend to report more	Cosner (2011) students in each of the three project schools made gains in standardised literacy tests over the three years of the study (USA).	Copland (2003) Schools were categorised as novice, intermediate or advanced in implementation of the cycle of inquiry. At schools in the advanced range of inquiry, teacher communities engaged multiple-level inquiry cycles that explicitly addressed connections and gaps across the school system, classroom practice, and student	Croxford et al. (2009) Describes QAE system in Scotland and, in particular, the expecations for self evaluation and the local education authority role in supporting this. Questions the extent to which a top down approach, in which performance goals are set by the inspectorate can be compatible with the school improvement goals of teachers. The authors suggest that the model may encourage performativity and compliance rather than improvement (Scotland.non-empirical).

		involvement (parents, pupils etc) (England).	specific activities and better explicit improvement goals ; whilst the schools with very low improvement are characterized by a stereotypical tendency in SIP elaboration.' (p92)		outcomes. Schools that stayed in the novice group treated the cycle of inquiry as a compliance exercise and teacher isolation and hierarchical leadership were barriers to change. Schools that moved into the intermediate group learned to value the process of inquiry and its potential for dealing collectively with problems in the school and improving teaching and learning (USA).	
	Education Scotland (2012) The use of self- evaluation to drive improvement is noted overall as an 'aspect for improvement'. They conclude 'approaches to self-evaluation need to impact on young people's learning and achievements, including their attainment' (Scotland).	**Ehren et al (2014)** inspection frameworks that set expectations about standards in education can drive school evaluation and the capacity of schools to improve (The Netherlands).	**Cowan (2008)** impact evidence for SSE including: improved planning; understanding of performance; more focused accountability among staff for performance; focus on pupil outcomes; introduction and evaluation of targeted interventions (England).	**Cowan (2008)** Over the two years, improvement in standards was found in many schools, particularly for groups that had been identified as underachieving through the more robust SSE process. (England).	**Dembosky et al (2006)** Teachers use data to adjust classroom instruction in three ways: whole class instruction, group instruction (most prevalent), and individualized instruction (USA).	**Davies and Rudd (2001)** staff felt overworked (England).

	Estyn (2012) In approximately one fifth of schools data was not being used effectively to improve schools, training and support for these schools is recommended (Wales).	Gray et al(2011) In all countries, teachers believed that Internal QAE was most likely to improve quality, citing teacher and school self-evaluation, analysis and tracking of pupil progress and target setting by the school and that these practices had most influence on them (Denmark, England, Finland, Scotland, Sweden).	Demie (2003) Schools use performance data and research findings effectively for school improvement purposes (England).	Demetriou and Kyriakides (2012) all three groups of schools that had implemented self- evaluation saw greater student progress in performance on a maths test than the control group. Schools that had been introduced to the dynamic model which links evaluation priorities to school effectiveness factors did significantly better than the other groups (Cyprus).	Farrell (2014) Some of the changes made to teaching were: targeting borderline students, tracking students, re- teaching those items where students had not understood a concept, influencing vocabulary use in lessons (USA).	Hall and Noyes (2007). In 'Centralised' schools, senior leaders made unannounced classroom observations and made judgements against inspection criteria, middle leaders felt pressured to follow up on these judgements with colleagues, with feeling of upset. 'Resisting' schools viewed self evaluation as a bureaucratic exercise to 'jump through hoops' and unrelated to school improvement. Teachers who express different views, for example of the importance of individual teacher autonomy, may be considered out of date or incompetent. Curriculum leaders are increasingly expected to monitor the work of their teams and some experience difficulties in combining this with supportive relationships.. (England).

	Karagiorgi (2012) Teachers found difficulties in actually introducing the intervention planned as a result of SSE, reporting a lack of time to do so and the timely availability of resources (Cyprus).	**Karagiorgi et al (2015)** The SSE project was successful in terms of self- reported changes made within the school, with schools valuing the autonomy in determining goals and actions (Cyprus).	**Education Scotland (2012)** note examples of good practice of use of self- evaluation to support school improvement in some schools (Scotland).	**Dunn et al (2013)** Student achievement and progress in the district was higher, and improving more rapidly, than elsewhere in the state. Although the authors state that the improvement cannot be linked directly to the systematic and supported use of data to improve learning, it is likely that this is a contributory factor (Canada)	**Gallimore et al (2009)** Seeing causal connections fosters acquisition of key teaching skills and knowledge, such as identifying student needs, formulating instructional plans, and using evidence to refine instruction (USA).	**Hartong (2012)** She criticises the framework for the way in which it narrowly defines success and improvement, with improvement measured through subsequent SEIS results and argues that it reduces school and teacher autonomy and self-conception by offering a rigid model for what it means to be 'good'(Germany).
	Quint et al (2008) Although schools found the data coaches helpful, gains in reading were no different from in comparator schools. They argue that because of the context of increasing data use in all schools in the area, results are inconclusive regarding the data coach role (USA).	**MacBeath (2008)** In a system in which SSE is externally verified and supported by External School Review there was Increased sharing and dialogue about classroom practice; a move to a more rigorous and systematic way of assessing practice. SSE is building capacity for a shift to a mindset of organizational learning (Hong Kong).	**Ehren et al (2014)** inspection frameworks that set expectations about standards in education can drive school evaluation and the capacity of schools to improve. Changes in self-evaluation lead to improvements in capacity building which leads to improvements in school effectiveness (The Netherlands).	**Gallimore et al (2009)** Grade level teams using an inquiry-focused protocol to solve instructional problems significantly increased achievement (USA).	**Halverson et al. (2007)** there was ongoing dialogue between teachers and leaders on the use and implications of data which led to improvements in instruction (USA).	**MacBeath (2008)** Other consequences included a significant period in which many staff reported high levels of stress and workload and anxiety. Leadership steering on this issue strongly mediated levels of stress among staff, so where ESR was seen as an opportunity rather than a threat, this was very helpful (Hong Kong).

	Timperley and Parr (2009) Compares the impact on student achievement of two initiatives to support schools in using data for instructional improvement in literacy. In the first study, school participants attended a training course on data use, with no impact on student achievement (New Zealand)	McNamara et al. (2011) Four primary schools in Iceland which had developed their own process and indicators with the support of two researcher-consultants had been successful in embedding a culture of self-evaluation, reflection and improvement within the school (Iceland).	Estyn (2012) Schools were found to use a range of strategies for improvement. These strategies include targeted support for groups of pupils, revising curriculum content or organisation, and increased professional learning. Estyn found that, on the whole, the impact of improvement strategies was good, with evidence of year-on-year improvement. Local authority staff were using the data to inform their support work and governors were using the data sets to challenge schools (Wales).	Lai and McNaughton (2013) Interventions significantly improved student achievement over 3 years, and these achievement gains were sustained after the interventions (New Zealand).	Hartong (2012) Classroom practice has been sustainably transformed through training in systematic teaching development, which belonged to the most important agenda settings in almost every analysed school' (p756) (Germany).	Simons (2013) Prescription of framework can lead to cynicism about the process... Teacher resistance may prevent top-down initiatives from being successful, especially where unions are strong (e.g. Ireland) or teachers' motivation may be weak (e.g. in Spain where teachers are civil servants for life and where there is no institutional leadership structure to promote and support SSE) Increased test-based accountability may mean that other factors and achievements are not given attention (international).
	Vazquez and Gairin (2014) Studies (optional) self-evaluation using a tool developed from the work of the authors. The research showed that, in general, conditions were in place for self-evaluation and planning for improvement to take place, although they believe that settings need more support in	Neil and Johnston (2005) Whole school self-evaluation can extend the use of school specific data as an aid to critical reflection and evaluation (Northern Ireland).	Hall and Noyes (2007) Three categories of schools: 1. Collaborative – where self-evaluation was seen as contributing to school improvement as a shared endeavour involving middle leaders and teachers (England).	Marsh et al. (2010) Studied the impact of data coaches working with teachers on analysis of data and implications for instruction. Data analysis support has a significant association with both perceived improvements in teaching and higher student achievement (USA).	Hofman et al (2010) Schools with effective SSE received high scores on quality of teaching when inspected by the Dutch inspectorate (The Netherlands).	Vazquez and Gairin (2014) 'Aspects like resistance to change and evaluation (fear of evaluating and being evaluated), the high mobility of their human resources and the culture of orality as the main means of transmitting knowledge work against processes that require permanence over time, collective reflection, recording and systematization of new knowledge, revision of 'what has been done' and

	doing so. However, actions resulting from the self-evaluation were found to be highly vulnerable and not sustained (Uruguay).					projection of 'what should be done'. The analysis also reveals a heavy imprint that associates self-evaluation practices with control, stigma and sanction.' (p 341)(Uruguay).
		Nevo (2001) SSE can help promote 'evaluative literacy' making school personnel more sensitive to areas in need of improvement; SSE can help to extend the scope of external evaluation by pointing out local opportunities for development and also factors relevant to the challenge of the school (Israel, non-empirical).	**Karagiorgi et al. (2015)** Schools identified specific improvement actions based on analysis of the self- evaluation results. For example one school decided to implement classroom visits and peer observation. Schools then organised professional development, usually involving external partners, to support improvement actions and monitored implementation. In the final stage, the external facilitator met with the school team to review outcomes and revise action plans for further improvement (Cyprus).	**McNaughton et al (2012)** Replication study of model described by Lai et al (2009) with similar results of increased and sustained student achievement (2012).	**Macbeath (2008)** teaching was more engaging and learner centred (Hong Kong).	**Wong and Li (2010)** Negative effects reported by all three schools were those of workload and stress associated with the inspection. Comments on workload were greatest from the weakest setting (which had problems with recruiting and retaining suitable staff) (Hong Kong)

		O'Brien et al (2015) Capacity was built, expectations were raised and the service became more professional following self evaluation (Ireland).	Leung (2005) The case study school is found to be highly successful in employing SSE for school development and accountability (Hong Kong)	Stringfield et al (2008)Longitudinal analyses of outcome data from 12 Welsh secondary schools indicated that 4 years after the High Reliability Schools project was initiated, student outcomes at the sites were strongly positive. Additional quantitative and qualitative data, gathered 5 years after the end of the intervention, indicated that the majority of the schools continued using the high reliability principles and continued making strong academic progress (Wales).	Marsh et al. (2010) Studied the impact of data coaches working with teachers on analysis of data and implications for instruction. Data analysis support has a significant association with both perceived improvements in teaching and higher student achievement (USA).	Wroe and Halsall (2001) Data was seen to be dangerous as potentially swamping the school and further confusing the issue. Processes derived internally through self-evaluation were seen to be more useful in generating improvements in practice. Quantitative 'target setting' was seen as less useful and potentially leading to teaching to the test (England).

		Ritchie (2002) The self-evaluation which included ofsted criteria for lesson observations, helped make the school 'self-inspecting' and they would have a wealth of evidence to show Ofsted when they next came to visit (England).	Neil and Johnston (2005) Whole school self-evaluation can contribute to school improvement...extend existing methods of evaluating teaching and learning; provide an information base for school development planning and its effective implementation; identify priorities for the professional development of teachers in the school; capitalise on schools' discretion as to the agenda for improvement and in-depth staff ownership of its direction and detail; contribute to a framework for managing change in schools; develop a perspective among teachers which goes beyond their own classroom(Northern Ireland)	Timperley and Parr (2009) Second project with data use supported by needs analysis and facilitators working with professional learning communities. Student achievement gains beyond what might have been expected were observed in all schools in the project (New Zealand).	McNaughton et al (2012) It was the process of using data to develop context-specific teaching programmes and its concomitant operationalizing of instructional content which was demonstrated to be effective rather than the specifics of approaches to instruction (New Zealand).	

		Schildkamp et al (2012) Although self-evaluation results may not lead directly to improvement actions they may influence the respondents' thinking and influence the direction of school improvement (Flanders, The Netherlands)	**Schildkamp et al (2012)** Schools which were judged to use SSE to promote school improvement took measures to improve their schools' quality of education (Flanders, The Netherlands).		**Timperley and Parr (2009)** Second project with data use supported by needs analysis and faciliators working with professional learning communities. Classroom observation by the researchers provided evidence of changed instructional practice and teacher content knowledge (New Zealand).	
		Simons (2013) defines SSE as an exercise which provides 'insight into the educational experiences of students' (p2)as more than those measured by test data and required for externa, accountability purposes (international, non-empirical).	**Supovitz and Klein (2003)** Schools used data for:Informing instruction; developing assistance plansfor low-performing students; planning professional development; setting goals; motivating faculty and students, visually stating school priorities and goals, and communicating with parents (USA).		**Wayman and Stringfield (2006)** In three schools that used data effectively there was increased sense of teacher efficiency, better response to student needs, reflection on practice and collaboration (USA).	

		Yeung (2011) System in which self evaluation against a prescribed framework is followed by validation through external review in a 3-4 year cycle. 85% of teachers agreed that it could 'help schools to understand own weaknesses, strengths and identify needs for improvement' and 78% indicated that their schools had made changes as a result of the process (Hong Kong).	**Wohlstetter et al (2008)** school systems required school improvement plans that measured progress towards goals set in internal evaluation (USA)		**Webb et al (1998)** Finnish SSE with non-hierarchical, small and family atmosphere in schools was much less rigorous and systematic than inspection in England but led to changes in classroom; pupil and parents feedback helped make changes and encouraged pupil independence (Finland and England).	
					Wroe and Halsall (2001) Reflective exercises with students and setting targets with clear learning outcomes discussed in appraisal meetings, led to improvements in pedagogy and learning. Processes derived internally through self-evaluation were seen to be more useful than test data in generating improvements in practice with	

				quantitative 'target setting' was seen as less useful and potentially leading to teaching to the test (England).	

Table 2. Studies reporting on conditions of effective internal evaluations

	conditions					
	evaluation literacy	resources	leadership	external support	supportive culture	accountability
Country	international (2 studies), England/ Germany/ Lithuania/ the Netherlands/ Poland (1 study), England/ Scotland/ the Netherlands/ Northern Ireland,/Denmark and Belgium/ Germany (Hesse and Lower Saxony) (1 study), England (4 studies), ireland (2 studies), Italy, New Zealand (5 studies), the Netherlands (4 studies), USA (13 studies), Wales (2 studies).	international (1 study), South Africa/New Zealand/England/the Netherlands/ Belgium- Flanders (1 study),England/ Germany/ Lithuania/ the Netherlands/ Poland (1 study), The Netherlands/Belgium-Flanders, Cyprus, England (3 studies), Germany, Hong Kong, New Zealand, The Netherlands, USA (13 studies).	England/ Germany/ Lithuania/ the Netherlands/ Poland (1 study), Belgium-Flanders (2 studies), Canada (2 studies), Cyprus, England (3 studies), Hong Kong, Mauritius, Northern Ireland, Norway, the Netherlands (3 studies), the Netherlands/Belgium-Flanders, USA (12 studies).	England/ Scotland/ the Netherlands/ Northern Ireland/ Denmark and Belgium/ Germany (Hesse and Lower Saxony)(1 study), Belgium- Flanders, Canada, Cyprus, England (4 studies), Europe, Germany, Iceland, Ireland, Ireland and Iceland, New Zealand (2 studies), Norway, The Netherlands, USA (7 studies), USA and UK.	England/ Germany/ Lithuania/ the Netherlands/ Poland (1 study), Belgium-Flanders (2 studies), Chile, England (3 studies), Europe, Hong Kong (2 studies), Ireland, Italy, New Zealand, The Netherlands/Belgium-Flanders, USA (9 studies).	England/Scotland/Swede/ Denmark/ Finland (1 study), England/ Scotland/ the Netherlands/ Northern Ireland,/Denmark and Belgium/ Germany (Hesse and Lower Saxony) (1 study), England/ Germany/ Lithuania/ the Netherlands/ Poland (1 study), international (2 studies), Belgium/Flanders, Chile, England (2 studies), Hong Kong, the Netherlands, USA (4 studies).
Year in which studies were reported	2000, 2003, 2005, 2006, 2007 (3 studies), 2008 (8 studies), 2009, 2010 (4 studies), 2011, 2012 (3 studies), 2013 (5 studies), 2014 (6 studies).	2001, 2003(2 studies), 2005 (3 studies), 2006 (2 studies), 2007 (2 studies), 2008 (2 studies), 2010 (2 studies), 2012(3 studies), 2013 (2 studies), 2014 (5 studies), 2015.	2003 (3 studies), 2005 (4 studies), 2006 (2 studies), 2008 (2 studies), 2010 (3 studies), 2011 (3 studies), 2012 (2 studies), 2013 (2 studies), 2014 (6 studies), 2015 (2 studies)	2001 (2 studies), 2003 (3 studies), 2004 (2 studies), 2005, 2007 (2 studies), 2008 (3 studies), 2009, 2010 (2 studies), 2011, 2012 (2 studies), 2013 (3 studies), 2014, 2015 (2 studies)	2000, 2001, 2003 (3 studies), 2004, 2005 (2 studies), 2006, 2007, 2008 (4 studies), 2009, 2011, 2012 (2 studies), 2014 (4 studies), 2015	2000, 2004 (2 studies), 2007, 2008 (2 studies), 2011 (2 studies), 2013 (2 studies), 2014 (5 studies)

Summaries	Anderson et al. (2010) ... actions influencing data use by principals and teachers include modelling data-informed decision making, setting and monitoring expectations for data use, providing tools to assist with data collection and interpretation, and providing or developing expertise to support data use at the school level (USA)	Cosner (2011) The development of knowledge about student learning and of using this to improve instruction took time for the teams in this study (USA)	Ah Teck and Starr (2014) found that principals were resistant to a formalised quality assurance process, citing lack of expertise in collecting and analysing data, the time it would take up (Mauritius).	Ancess, Barnett and Allen (2007) report on two cases of a university working with school partnerships and providing support. The authors conclude that the cases demonstrate the critical importance of researcher–practitioner collaboration and research partners' responsiveness to the identified goals and emerging needs of partner organizations and of the schools and practitioners with whom they work (USA).	Bubb and Earley (2008) It was essential that staff were convinced that change was both necessary and possible and that appropriate support and professional development were in place (England).	Ehren et al. (2014) inspection frameworks that set expectations about standards in education can drive school self- evaluation and the capacity of a school to improve. Schools that were high on setting expectations, accepting feedback, and stakeholders' sensitivity were also more likely to narrow the teaching strategies and curriculum; send a more 'rosy' picture of the school in documents sent to the inspectorate... (The Netherlands).

Barrett(2009) She also characterizes the talk about students as frequently superficial, laden with stereotypes, and focused on explanations for student failure that reside outside the control of the teacher. The presence of a facilitator and the availability of tools for displaying and reviewing data appeared to have little effect on ingrained ways of classifying students according to perceived effort, motivation, and ability (USA).	**Copland (2003)** Principals maintain the vision for change through hiring committed staff and firing those that were resistant and through supporting inquiry, for example, by the provision of dedicated time for collaboration (USA)	**Anderson et al. (2010)** Principals often aware of conditions that fostered or inhibited data use, but a majority of principals appeared to have externalized those conditions. Few indicated taking action to improve most of the conditions. Some conditions influencing data use by principals and teachers may be more accessible to principal influence (e.g., teacher capacity for data use, time for data use) than others (e.g., timeliness of data, quality of data). A minority of principals take action to shape those school-level conditions most open to their influence (USA).	**Anderson et al. (2010)** Principal use and support for teacher use of data is strongly shaped by the actions of district office leaders in the context of varying state accountability requirements and support (USA).	**Caputo and Rastelli (2014)** schools with low achievement improvement level mainly attribute responsibility for student education to the social context (Test Value = 2.18); whilst schools with high SES perceive themselves effective in improving education (Italy)	**Hall and Noyes (2007)**. In 'Centralised' schools, senior leaders made unannounced classroom observations and made judgements against inspection criteria, middle leaders felt pressured to follow up on these judgements with colleagues, with feeling of upset. 'Resisting' schools viewed self evaluation as a bureaucratic exercise to 'jump through hoops' and unrelated to school improvement. Teachers who express different views, for example of the importance of individual teacher autonomy, may be considered out of date or incompetent. Curriculum leaders are increasingly expected to monitor the work of their teams and some experience difficulties in combining this with supportive relationships.. (England).

		Cowan (2008) Improved data availability helped schools undertake sharper data analysis which in turn supported better assessment of performance and identification of action required to address underperformance (England, non-empirical)	Bubb and Earley (2008) Strong leadership was found to be essential in ensuring impact on teaching quality and/or student outcomes (England)	Cowan (2008) The rigour of the SSE was supported by the School Improvement Partner (England).	Chapman (2000) The environment of trust was seen as vital in peer observation (England).	Honig and Ikemoto (2008) The research team concluded that factors contributing to the effective use of data for school improvement are: Accountability incentives to use data; Teacher flexibility to change instruction (USA)
	BERA (2014) Teachers need […] to be equipped to engage in enquiry-oriented practice – having the capacity, motivation and opportunity to use research-related skills to investigate what is working well and what isn't fully effective in their own practice. Using data from a number of different sources teachers can identify problems together with interventions which are likely to lead to improvements. (p 30)(UK).	Davies and Rudd (2001) Time for development activities were seen as essential (England). Davies and Rudd (2001) SSE packages could help schools evaluate (England).	Copland (2003) Formal leaders frequently act as catalysts for change in initial stages (of cycle of inquiry process) with distributed structures emerging as the model develops. Principals maintain the vision for change through hiring committed staff and firing those that were resistant and through supporting inquiry, for example, by the provision of dedicated time for collaboration.(USA).	Davidsdottir and Lisi (2007) researchers coached school-evaluation teams in their self-evaluation efforts and assessed change in the schools. The schools received empowerment-based support, where the researchers taught staff to evaluate school work and take responsibility for development based on outcomes from longitudinal data collection and dissemination of evaluation information (Iceland).	Copland (2003) Schools that stayed in the novice group treated the cycle of inquiry as a compliance exercise and teacher isolation and hierarchical leadership were barriers to change (USA)	Janssens and Van Amelsvoor (2008) Where the SSE largely matched the inspection framework, as in Denmark, England, Scotland, Northern Ireland, this increased the accountability orientations ... Where the national system of inspection has an external evaluation aspect, the SSE occupies a stronger position. (Denmark, the Netherlands, Belgium, Germany, England, Northern Ireland and Scotland).

	Blok (2008) SSEs were of low quality, often failing to answer questions set at the onset of the process. They conclude that considerable support and guidance was needed for schools in completing SSEs (The Netherlands).	**Dembosky et al (2006)** Several schools and districts lack sufficient technology to make the best use of data. Across all case study districts, teachers and principals identified lack of time as the biggest constraint on their ability to review and reflect on data. Many teachers expressed a desire to learn more about how to analyze and interpret data but availability of training was insufficient (USA).	**Cosner (2011)** Principals, whose roles also evolved over time, were important in the design and introduction of tools and processes that supported grade-level data-based collaboration that in turn influenced the ways in which student learning knowledge and instructional considerations developed from data-based practices (USA).	**Davies and Rudd (2001)** LEA advisers or colleagues helped in the role as a critical friend (England).	**Earley and Bubb (2014)** schools need to have a supportive culture (England, non- empirical)	**Jimerson (2014)** There was a belief that data use was primarily for accountability purposes. Even when teachers wanted to improve, 'the shadow of testing and accountability loomed large when the term 'data use' is part of the conversation (p 10) (USA).
	Caputo and Rastelli (2014) evaluated a training programme based on a self-evaluation. The training programme helps teachers formulate action goals in a school improvement plan, support for professional development and opportunities for working in professional learning communities (Italy).	**Earley and Bubb (2014)** schools need to resources, particularly of time (England, non-empirical)	**Demie (2003)** The leadership of the headteachers is key in raising the level of expectations (England).	**Demie (2003)** Training and support from the LEA contributes to effective use of data (England).	**Gallimore et al (2009)** positive outcomes are more likely when teams are teaching similar content, led by a trained peer-facilitator, using an inquiry focused protocol, and have stable settings in which to engage in continuous improvement (USA).	**Macbeath (2004)** Argues for a negotiated approach to self- evaluation in which external standards, inspection and challenge have a place in supporting challenge but the need for capacity building to inform transformational change is essential (England, non- empirical)

Chapman (2000)	Farrell (2014)	Devos and Verhoeven (2003)	Devos and Verhoeven (2003)	Herman and Gribbons (2001)	Montecinos et al(2014)
Self-evaluation can be lacking in that teachers do not always have the necessary expertise; may not be able to identify needs or may not have the necessary challenge in their collaborative partnerships within school (England).	Support for data use was also varied, including support for teacher collaboration, coaching positions, professional development, training for new employees, data management systems and other tools and technology, scheduled time, practices and routines and rewards and incentives. The resources provided were found to influence the extent to which data was used to improve instruction (USA).	Change is dependent on internal factors, particularly on the leadership of the school (Flanders).	The external perspective provided by the researchers was helpful in identifying what would otherwise have been blindspots for self- evaluators (Flanders).	"The importance of trust and efficacy were underscored. Combating a siege mentality and getting beyond blame and defensiveness to action are problems that go far beyond technical and mundane aspects of data use." (p 18)(USA)	Teachers' responses to self evaluation reflected their perceptions of ... internal accountability. In one of the schools they describe a positive response by teachers who valued the opportunity to contribute to self evaluation and identify how to improve. One other school viewed the self- evaluation process as a need to demonstrate competence to external evaluators. The teachers in the remaining four schools completed the tasks set by administrators in their school but were cynical about the extent to which their contributions were taken into account, compared with the opinions of external evaluators (which they considered likely to be focused on fault finding). (Chile).

	Coburn and Talbert (2006) They advocate the use of varied sources of evidence and leadership from the district in mediating understandings of evidence based practice. They also advocated the importance of enabling communication across all levels of the system to share understandings and practice (USA)	Hartong (2012) She criticises the framework for the way in which it narrowly defines success and improvement, with improvement measured through subsequent SEIS results and argues that it reduces school and teacher autonomy and self-conception by offering a rigid model for what it means to be 'good'(Germany).	Dunn et al (2013) Principals were expected to lead and facilitators were employed to support professional learning communities in using data for enquiry. Principals in the district were responsible for promoting collaborative discussion and learning around the data collected and for monitoring, supported by critical friends, the changes in classroom practice (Canada).	Dunn et al. (2013) Systems for data management were introduced and supported by workshops and professional learning communities within schools. Professional learning communities had the support of a trained facilitator and principals were responsible for creating conditions for collaboration within their schools, as well as for monitoring with a critical friend, the impact on classroom practice (Canada).	Honig and Ikemoto (2008) The research team concluded that factors contributing to the effective use of data for school improvement are: Accountability incentives to use data; Teacher flexibility to change instruction (USA)	Ozga et al (2011) Overall, the authors comment on tensions between what they term as 'hard' governance, composed of externally determined regulation, benchmarks and targets and 'soft' governance whereby a cycle of self-evaluation and improvement actions leads to a continuous process of educational improvement. (England, Scotland, Sweden, Denmark, Finland)
	Cosner (2011) need for development of knowledge about student learning and of using this to improve instruction (USA)	Hofmann et al (2005) propose with a framework (p 261) for comparing SSE instruments according to an accountability, a school improvement objective and for reliability and validity. They propose six uses of the framework in judging instruments for: reliability; scientific standards and validity; usefulness and standards for evaluation; cover of	Earl (2008) Although teachers had sufficient trust to share and review their student data, skilled leadership was required to use those data to investigate their implications for teaching practice. The principal used data charts to re-focus the discussion on data, with teachers tending to deviate into more general issues of teaching practice or factors that affect learning (USA).	Farrell (2014) Support for data use was also varied, including support for teacher collaboration, coaching positions, professional development, training for new employees, data management systems and other tools and technology, scheduled time, practices and routines and rewards and incentives. The resources provided were found to influence the extent to which data was used to improve instruction (USA).	Lachat and Smith (2005) Schools had a data coach and used data teams, both of which contributed to engagement, with all schools committed to data use to inform school improvement by the end of the study (USA).	Rallis and MacMullen (2000) suggest that both external and internal accountability are needed to drive improvements within inquiry minded schools being the key to combining the two (USA).

		quality dimensions at different educational levels; suggestions and tools for school improvement; focus or purpose of the instrument (The Netherlands).				
	Earley and Bubb (2014) schools need research literate staff (England, non-empirical)	**Kallemeyn (2014)** The principal of a school with strengths in data use directed time for teams, including instructional coaches, to meet to review data and use it to plan instruction (USA).	**Earley and Bubb (2014)** Leadership is required to create and embed an inquiry- oriented culture and to provide the resources, particularly of time, to allow staff to collaborate, agree common goals for inquiry and discuss findings and to put support in place for staff to develop research expertise. (England, non- empirical)	**Gallimore et al (2009)** positive outcomes are more likely when teams are led by a trained peer- facilitator (2009)	**Leung (2005)** In a school which had been successfully using SSE since 1997 essential conditions were a dynamic pattern of shared values which is open to constant revision and changes; a genuine belief in the talents of every single individual at the school (Hong Kong).	**Schildkamp et al (2014)** Researchers noted that pressures from the accountability systems influenced the use of data for decision making (England, Germany, Lithuania, the Netherlands, Poland)

Estyn (2012) Approximately, one-fifth of schools were not using the data effectively to plan for improvement and Estyn recommend that training and support is provided for schools in both interpreting the data and using to improve (Wales).	**Karagiorgi (2012)** teachers said that they had insufficient time to introduce planned improvements following a process of self-evaluation (Cyprus).	**Ehren et al. (2014)** Principals have a key role in mediating external accountability policies. Their interpretation of accountability standards and policies is framed by the context in which they function; their beliefs, histories, and agenda, and their interpretation of these standards will influence their responses to inspection standards and how they will adapt school policy and structure to meet the standards (The Netherlands).	**Hartong (2012)** This research considers the use of a school self-evaluation tool (SEIS) developed by the Bertelsmann Foundation and its use in Lower Saxony. The tool collects data from teacher, student, parent and principal questionnaires and focuses at school rather than class level. Participating schools are required to share their data with their networks. SEIS is introduced in schools via training and its use is supported by consultants and coaches (Germany).	**MacBeath (2008)** SSE teams were most effective when they were afforded sufficient scope to use initiative and had ownership of the process; when they were a cross section of staff including all levels of seniority (Hong Kong).	**Schildkamp, Lai and Earl (2013)** data use needs teacher collaboration, ownership and autonomy. Data use is underpinned by policy with need for a balance between data use for accountability and support for its use for improvement (international, non-empirical)
Geijsel et al (2010) Lack of data literacy proved a barrier. The culture of learning from the data needed to be spread throughout the school (The Netherlands)	**Lachat and Smith (2005)** Leadership and support, including a time allocation, for collaborative working were found to be critical to success (USA).	**Emstad (2011)** It is the behaviour of the principal that is the key factor in both prioritising the activity above other initiatives in the school and in allowing time and conditions for discussion and reflection across the school (Norway).	**Herman and Gribbons (2001)** Research uses data from partnerships between a university and teams in two groups of schools. The researchers initially worked with school teams on data analysis, using district longitudinal data and any other data available, to produce a report which summarised 'how well are we doing?' 'are	**Marsh and Farrell (2015)** Where trust and a history of working together was established, data use to improve instruction development more rapidly.	**Simons (2013)** External imperative is required to get SSE underway, with voluntary SSE only working in committed schools (adoption in for example, Norway, was limited while it was optional)Mandatory SSE is only effective if schools are convinced of its usefulness to enhance their work in teaching and learning. Otherwise they see it as too time-consuming and as a

			we well serving all subgroups?' (USA)		distraction(international, non- empirical).
	Hall and Noyes (2007). 'Collaborative schools built shared knowledge and analyses through classroom observation, discussion and interpretation of data (England).	**Lai and Hsiao (2014)** found that barriers to effective use of data for school improvement are: Use and maintenance of data management systems, including ensuring ongoing quality assurance of data in the system; Mismatches between different systems preventing them being used efficiently in combination (New Zealand)	**Farley- Ripple and Buttram (2014)** In schools that were using professional learning communities to identify where and how instruction might be improvement, there was a clear vision at school and district level for the improvement process, school leaders were actively engaged. In the other schools, where the support and drive from both district and school leaders was lacking, there was less evidence for the impact of collaborative data use (USA).	**Hofman et al (2010)** Key factors for effective SSE that independent external validation is sought for internal judgements on quality; presenting information about quality to stakeholders, both external (e.g. parents, local authorities) and internal (staff, students) (The Netherlands).	**McNamara and O'Hara (2006)** internal self-evaluation conducted by teachers was seen very positively, with evidence that without the risk of criticism from outside, teachers were prepared to challenge each other and to use the outcomes constructively (Ireland).	**Supovitz and Weathers (2004)** An evaluation system where peer principals and district officers monitor specific local policies across the district, with each school getting one or two visits per year with constructive feedback to the school created a sense of accountability and urgency. But, despite anonymity, principals felt that snapshots (other principals and district officers monitoring their schools with constructive feedback) were used to judge their school's performance (USA)

Hofman et al (2010) key factor for successful SSE is that schools are in touch with their pupils (the Netherlands).	**Leung (2005)** In a school which had been successfully using SSE since 1997 an essential conditions is a comprehensive SSE system which is an on-going cycle of planning, trying, and evaluating (Hong Kong).	**Geijsel et al (2010)** School leaders need to share an aim of learning from the data and that trust needs to be established, so that 'principals … understand that the scores were not being used to judge the school but to initiate processes' (p68)(The Netherlands).	**Honig and Ikemoto (2008)**External support personnel contributing by providing professional development, access to research on learning, fostering joint working, developing tools, e..g. for learning walks, with all support adapted to local needs and working alongside educators (USA).	**Meuret and Morlaix (2003)** The main determinant for a positive attitude to self-evaluation was a belief in the impact of the process on school effectiveness and ability to improve. The participatory nature of the process is essential in its effectiveness (Europe).	**Vanhoof et al(2014)** External accountability expectations link with data use, however from the perspective of school development, data act as a guide in taking decisions for policy and practice (Flanders)
Honig and Ikemoto (2008) Factors contributing to effective use of data for school improvement include: Accessibility and timeliness of data; Validity of data; Staff capacity and support provided (USA).	**Marsh and Farrell (2015)** conclude that leadership, time, norms, supporting artefacts, trust required to build capacity for data use … lack of time could prevent response to some identified needs (USA).	**Kallemeyn (2014)** The principal led and modelled a school wide cycle of inquiry for improvement (USA).	**Janssens and Van Amelsvoor (2008)** Where a range of providers of training were available (and not just centrally, or by the inspection service itself), this was seen as favourable to improvement orientation (England, Scotland, the Netherlands, Northern Ireland, Denmark, Belgium, Germany).	**Montecinos et al(2014)** Teachers' responses to self evaluation reflected their perceptions of the culture of the school, including internal accountability, internal structures for teacher participation in decision making and social capital (Chile).	**Yeung (2011)** survey and questionnaire data from primary school curriculum leaders found that most believed that the primary purpose of the evaluation system is for monitoring effectiveness, with more weight given to external reviewers judgements than those from self review (Hong Kong)

Janssens and Van Amelsvoort (2008) A key aspect is the amount and type of steering given to schools about the SSE process and framework (England, Scotland, the Netherlands, Northern Ireland, Denmark, Belgium, Germany).	**Means et al. (2010)** Effective data-informed decision making requires not only access to useful data but also well-designed supports such as leadership to model data use, technical support for data interpretation and supported time for reflection on data (USA, non-empirical).	**Karagiorgi et al (2015)** Support from the head had been essential (Cyprus).	**Karagiorgi et al (2015)** External support from the facilitator had been essential (Cyprus).	**Ryan et al (2007)** Barrier to evaluation was the bureaucratic and hierarchical organisation culture, which sometimes clashed with the horizontal leadership and open dialogue needed for evaluation (USA).	
Jimerson (2014) recommends that school leaders consider how professional development can be used to help teachers move towards a broader understanding of data use and of how it can be used to support school improvement with a balance between this and accountability concerns (USA).	**Ryan et al (2007)** insufficient training and lack of time to conduct evaluations were barriers (USA).	**Knapp and Feldman (2012)** In the schools studied, judged by student outcomes to be 'making progress', principals were able to identify commonality between external accountability demands and their own beliefs. There mediated external expectations to align with a vision for education in their schools which they encouraged staff to share, and developed internal quality mechanisms, linked to, but not limited to, requirements for teacher evaluation and test targets. They integrated information from student	**Lai and McNaughton (2013)** discussion of collected data in groups facilitated by an external researcher was important (New Zealand).	**Schildkamp et al (2012)** Attitudes of teachers and leaders towards self-evaluation and willingness to innovate were found to be strongly related to use of results for improvement (Flanders, The Netherlands)	

			and staff surveys, internal enquiry teams and external review to guide staff development, led on data use promote school improvement and modelled and promoted collaborative working (USA)			

	Lachat and Smith (2005) Although initially based on dialogue around quantitative data, when the high school teams collaboratively developed clearly focused questions, it helped them look beyond the data to examine other pertinent information, and they were far more likely to understand what the data meant for school improvement (USA).	**Schildkamp et al (2014)** Barriers to using data included lack of training, lack of time, lack of supporting systems for analysis. (England, Germany, Lithuania, the Netherlands, Poland)	**Lachat and Smith (2005)** Leadership and support, including a time allocation, for collaborative working were found to be critical to success (USA)	**McNamara et al. (2011)** Four primary schools in Iceland which had developed their own process and indicators with the support of two researcher-consultants had been successful in embedding a culture of self-evaluation, reflection and improvement within the school (Iceland).	**Schildkamp et al (2014)** Where teacher collaboration was common, as in England, Lithuania and Poland, interviewees were able to provide more concrete examples of the forums for collaboration (e.g. in subject department meetings) and of the impact of decisions made based on the data reviewed. In Germany and the Netherlands teacher collaboration was less commonly reported. (England, Germany, Lithuania, the Netherlands, Poland)	
	Lai and Hsiao (2014) About two-thirds of school clusters could produce high-quality self-evaluation data with support (New Zealand).	**Schildkamp, Lai and Earl (2013)** Enablers and barriers are time, data and data systems, multiple sources of data, data infrastructure (international, non-empirical).	**Leung (2005)** In a school which had been successfully using SSE since 1997 an essential condition is a combination of leadership by the school management team and distributed leadership (Hong Kong).	**Mc Naughton, Lai and Hsiao (2012)** replication of Lai et al (2009) working together in professional learning communities important (New Zealand).	**Supovitz and Klein (2003)** Barriers to use of data to improve instruction included scepticism and resistance from some teachers (USA)	

Lai and McNaughton (2013) The use of student achievement data is insufficient to lead to improvement, as the data must be linked to instructional practice, with data gathered through lesson observations, student work, student surveys etc. (New Zealand).	**Schildkamp et al (2012)** Time for reflection and discussion were also found to be significant for effective use of SSE (The Netherlands, Flanders).	**Marsh and Farrell (2015)** conclude that leadership, time, norms, supporting artefacts, trust required to build capacity for data use (USA).	**Meuret and Morlaix (2003)** The impact of the critical friend had been weak, despite the role being highly appreciated by respondents (Europe).	**Timperley (2008)** Teachers in the most improved school openly expressed uncertainty and sought help from others in the group (New Zealand).	
Lasky, Schaffer and Hopkins (2008) Scaffolds were used to structure meetings among teachers and principals about data. Participants initially focused on procedural elements and demonstrated lack of understanding. Attention needs to be focused on how to structure school activities, materials, and norms so that people can develop the 'inquiry habit of mind' (p105) and skills that allow for intentional, and possibly critical conversations anchored to student data that can inform teaching and	**Supovitz and Klein (2003)** Barriers to use of data to improve instruction included lack of time for data analysis (USA).	**Means et al (2010)** Leadership for data-informed decision making at the school level can extend beyond the principal. The case studies suggest that they may be performed by individuals in a variety of job roles; instructional coaches, department lead teachers, and instruction and assessment coaches were providing leadership for using data in many of the case study schools (USA)	**O'Brien et al (2015)** A facilitator led approach supports successful implementation of SSE (Ireland).	**Vanhoof et al (2011)** "attitude with regard to self evaluation", "self-evaluation as a policy action" and "self-evaluation as an act of research" are powerful predictors of the quality of self-evaluations (Flanders).	

	organizational improvement plans (USA).				
	Little and Curry (2008) A protocol for the discussion was regularly used at meetings of a 'critical friends' group'. The conversation provided insights into teaching and learning, but participants also showed a tendency to focus on procedure with some superficial examination of the evidence at hand (USA).	**Verhaege et al (2013)** Research examines purposes and mechanisms of five systems to support SSE. They point to differences between the systems and argue that schools need to be aware of a tool's purpose and mechanisms, as well as ownership of the data generated, before choosing it to support self evaluation (South Africa, New Zealand, England, the Netherlands, Flanders).	**Murnane (2005)** Schools needed leadership committed to data use for instructional improvement (USA).	**Sjobakken and Dobson (2013)** suggest that on the basis of this study the process of self-evaluation is an ongoing process where negotiated ownership developed over time is needed to secure sustainability (Norway).	**Vanhoof et al (2012)** Where principals had established a strong collaborative culture were more able to encourage teachers to use school performance feedback in a productive way (Flanders).

	Marsh et al. (2010) Studied the impact of data coaches working with teachers on analysis of data and implications for instruction. Data analysis support has a significant association with both perceived improvements in teaching and higher student achievement (USA).	**Wayman and Stringfield (2006)** Effective data use requires efficient data access, time to learn the system and examine student data (USA).	**Neil and Johnston (2005)** Support of leaders critical for whole school SSE (Northern Ireland).	**Supovitz and Weathers (2004)** An evaluation system where peer principals and district officers monitor specific local policies across the district, with each school getting one or two visits per year with constructive feedback to the school. Findings: Principals found results to be useful for gaining a clearer picture of the level of implementation in their own schools and were able to make amendments to strategies for improvement; a common language developed with which to engage with each other about reforms;Debriefing sessions between data collectors and the principal of a visited schools enabled professional conversations and shared learning; Maintaining reliability was a challenge as the data collection team expanded (USA)		

McNamara and O'Hara (2008) many countries have set expectations and provided frameworks for self- evaluation, but insufficient attention has been given to the need to support its use to improve schools. It is argued that enabling individual schools and teachers to self-evaluate effectively is a complex task that will require help and support from the community of professional evaluators (Ireland, non-empirical).	**Wayman et al (2007)**A study of data use in a school district found no integrated computer-based system for efficient data use (USA)	**Schildkamp et al (2014)**Supporting factors were clear vision and goals (England and Lithuania) and supportive school leaders (the Netherlands and Poland).(England, Germany, Lithuania, the Netherlands, Poland)	**Swaffield (2004)** critical friend can offer a different perspective, seeing things through a different lens, acting as a mirror and a sounding board, asking provocative questions, and in so doing contribute essential elements to school self-evaluation' (p275) (England, non-empirical)	**Schildkamp, Lai and Earl (2013)** data use needs teacher collaboration, ownership and autonomy. Data use is underpinned by policy with need for a balance between data use for accountability and support for its use for improvement (international, non-empirical)	

	McNamara and O' Hara (2012)	Wohlstetter et al (2008)	Schildkamp and Visscher (2013)	Swaffield and Macbeath (2005)	Schildkamp and Visscher (2013)	
	McNamara and O' Hara (2012) School leaders had no support from the inspectorate to generate evidence, guidelines unclear, no training or time available for internal evaluation, evalation documents often seen as compliance with rules. Lack of benchmarking data on student attainment, no consultation with pupils or parents, insufficient support, guidance and training meant its intention as a basis on which to conduct external evaluations was limited.	**Wohlstetter et al (2008)** For most school systems, taking the time and resources to develop specific goals geared toward their needs ended up being a pivotal aspect of using data purposefully. In one of the districts this took 3 years, but all teachers interviewed could articulate goals clearly (2008).	**Schildkamp and Visscher (2013)** Critical success factor is a school leader stimulating and facilitating the use of data. (The Netherlands, non- empirical)	**Swaffield and Macbeath (2005)** Different models of school self-evaluation put the critical friend in different roles, with responsibilities to other stakeholders. The critical friend as a supportive yet challenging facilitator, in a process tailored to the values of the school is one model, the priority for a critical friend actually working with a school evaluating itself using an inspection framework may be more of a regulator (international, non=empirical)	**Schildkamp and Visscher (2013)** Critical success factors are motivated staff; a school culture that is achievement-oriented; collaborative agreement on a clear set of goals and clear division of tasks between team members;pedagogical content knowledge (The Netherlands, non- empirical)	
	Mutch (2012) Suggests the need to build capacity for successful SSE (New Zealand, non-empirical).		**Schildkamp et al (2012)** Whether teachers used self- evaluation depended a lot on the school leader, with, for example, encouragement of the use of the tool and support for professional development found to be important.		**Simons (2013)** External imperative is required to get SSE underway, with voluntary SSE only working in committed schools (adoption in for example, Norway, was limited while it was optional) Mandatory SSE is only effective if schools are convinced of its usefulness to enhance their work in teaching and learning.	

					Otherwise they see it as too time-consuming and as a distraction (international, non-empirical).	
	Schildkamp et al (2014) Supporting factors were availability of expert advice on analysis and use of data (England and the Netherlands).(England, Germany, Lithuania, the Netherlands, Poland)		**Vanhoof and Van Petegem (2011)** pre-conditions for self-evaluation to be successful in influencing school improvement: The school team uses shared leadership as a means of creating involvement. (Flanders,, non- empirical)		**Vanhoof and Van Petegem (2011)** pre-conditions for self-evaluation to be successful in influencing school improvement: The school team is prepared to engage in systematic reflection, self- evaluation process is written into school policy(Flanders, non-empirical)	
	Schildkamp, Lai and Earl (2013) Enablers and barriers are culture of enquiry, training and support, disposition to use data, knowledge and skills (international, non-empirical).		**Wayman and Stringfield (2006)** Principals in a study of effective data use ensured that data were used for diagnostic purposes that were seen by teachers as beneficial and non- threatening (USA).		**Wohlstetter et al (2008)** leaders at all levels created an ethos of learning and continuous improvement rather than one of blame. Teachers relied on one another for support, new teaching strategies and discussions about data (USA)	

	Schildkamp and Visscher (2013) Critical success factors are a cycle of core activities; knowledge and skills for data use (The Netherlands, non-empirical)		**Young (2006)** Role of leaders important in providing vision, expectations and norms for collaborative discussion of data to improve instruction across teams. Where this was not in place, although teachers may have used data to inform their own practice, teachers worked individually with little collaboration (USA).		
	Simons (2013) training in SSE is helpful, but does not secure sustainability. Training is required in establishing criteria; setting boundaries for the evaluation; building on methods and skills that are already in place to support evaluation; analysing and making sense of data; political and interpersonal skills in sharing evaluation knowledge (International, non-empirical).				

	Stringfield et al (2008) A contributory factor in schools using the high reliability principles and continuing to make academic progress was the importance of data analysis and use to identify areas of improvement at all levels (Wales).				
	Supovitz and Klein (2003) Barriers to use of data to improve instruction included lack of technical expertise in analysing performance data (USA).				

	Timperley (2008) schools making sustained progress were characterized by regular, focused discussions on student progress data and implications for teaching approaches. Discussion among teachers were frequent, there was openness to challenge and teachers sought help from others. In schools making less progress, discussions were stuck in activity traps in which examining data and having conversations was seen as a good thing to do with only a vaguely defined purpose for doing so (New Zealand).				
	Timperley (2013) 'the capacity of schools to reflect on the quality and accuracy of their data and to perform accurate analyses relevant to their purpose is widely viewed as an integral part of effective self-review and evaluation' (p63). (New Zealand, non- empirical)				

	Trachtman (2007) through the use of a wide range of sources of evidence, the judgements against standards can be more nuanced and authentic compared to test based accountability applied elsewhere in the US system (USA, non-empirical)				
	Wayman et al (2007) A study of data use in a school district found no shared vision or understandings about what learning was and how data should be used to support teaching and learning; lack of definitions, protocols, or uniform procedures for data entry led to concerns about accuracy; principals and teachers were not adequately prepared to use data (USA).				

Table 3. Studies reporting on mechanisms of change of effective internal evaluations

	accepting and interpreting feedback	building capacity/organisational learning	implementing improvements
Country	Belgium - Flanders/The Netherlands (2 studies), Germany	Belgium- Flanders/the Netherlands (2 studies),Canada (2 studies) England (1 study), New Zealand (2 studies),the Netherlands (1 study), Scotland (1 study), USA (14 studies), Wales (1 study), England/Germany/ Lithuania/ the Netherlands/ Poland (1 study), international (1 study)	Belgium- Flanders, New Zealand, USA
Year in which studies were reported	2012 (2 studies), 2014	2001, 2004, 2005(2 studies), 2006 (1 studies), 2007 (2 studies), 2008 (4 studies), 2009 (2 studies), 2011, 2012, 2013 (4 studies), 2014 (4 studies), 2015.	2003, 2010, 2012
Summaries	**Gaertner (2014)** discusses on a system in Germany that provides student feedback to individual teachers rather than as a deliberate component of whole-school SSE, found that it is only when the feedback is shared and discussed that it contributes to organizational change. He suggests that student feedback needs to be integrated into a school-wide quality management system (Germany).	**Ancess, Barnett and Allen (2007)** A university working with a school partnership supported a working group by creating a structure, providing and reviewing data, setting up time to work together, helping schools to work through their planning and implementation processes, revising the structure, and providing and reviewing new data. In each cycle, learnings were documented and implemented by the organization; they were also used in developing the next cycle (USA).	**Anderson et al (2010)** Statistical data on a link between data use and student achievement was weak. Qualitative data suggested that it is the appropriateness of actions actually taken based on data-informed decisions that has an impact on the quality of teaching and learning (USA).
	Schildkamp et al (2012) Schools which were judged to use SSE to promote school improvement studied the feedback, discussed it and took measures to improve their schools' quality of education (Flanders, The Netherlands).	**Christie et al (2004)** describe a case of collaborative evaluation in a community college in California. An externally funded program was evaluated through the work of an evaluation team composed of external evaluators and faculty. As a result of the work of the learning community the team developed, attitudes to evaluation in the college changed and it impact on learning was valued. Evaluation capacity grew, with the external evaluator's role diminishing over time (USA).	**Devos and Verhoeven (2003)** following discussion with 'external experts' the schools agreed with the researchers about the implications of the results for their schools and on the changes that might be made to improve organisational climate. However, change occurred in none of the case study schools, due in two cases to the authoritarian leaders of the schools, who were either unable or unwilling to change. In the third school, it had been decided collaboratively that the proposed change would detract from the work of the school. The team conclude that the external perspective provided by the researchers was helpful in identifying what would otherwise have been blindspots for self-evaluators, but that change is dependent on internal

		factors, particularly on the leadership of the school (Flanders).
Vanhoof et al. (2012) investigated the use of pupil performance feedback, together with benchmarking data. Volunteer schools were provided with a support programme in interpreting and using data. However use of feedback was limited in planning school improvement actions, both in schools that had participated in the support programme and those which had not (Flanders).	**Cosner (2011)** Principals, whose roles also evolved over time, were important in the design and introduction of tools and processes that supported grade-level data-based collaboration that in turn influenced the ways in which student learning knowledge and instructional considerations developed from data-based practices (USA)	**McNaughton et al (2012)** it was the process of using data to develop context specific teaching and its concomitant operationalizing of instructional content which was demonstrated to be effective rather than the specifics of approaches to instruction (New Zealand).
	Dembosky et al (2006) Teachers collaborate in making decisions based on data. Most collaboration occurs within grade levels or, at the high school level, departments. In some schools data teams span grade levels. However, teachers frequently reported that they would collaborate more if they had more time. Some schools have built time into teachers' schedules specifically for reviewing student data and planning instruction, but this is rare (USA).	
	Dunn et al (2013) describes how the province of Ontario, Canada structured a large-scale initiative to ensure that all schools in the province had access to high-quality data and to develop a culture of inquiry, in which there is widespread capacity to work with data and where using data becomes a routine part of the operation of the educational system at all levels. Principals were expected to lead and facilitators were employed to support professional learning communities in using data for enquiry. Principals in the district were responsible for promoting collaborative discussion and learning around the data collected and for monitoring, supported by critical friends, the changes in classroom practice (Canada).	

		Earl (2008) Although teachers had sufficient trust to share and review their student data, skilled leadership was required to use those data to investigate their implications for teaching practice. The principal used data charts to re- focus the discussion on data, with teachers tending to deviate into more general issues of teaching practice or factors that affect learning (USA).	
		Farley- Ripple and Buttram (2014) A time allocation of 90 minutes per week was provided by the state for teachers to collaborate in discussion on using data for improvement. In only one of the case study districts had this resulted in positive change. The difference was that in this district there was drive and support, with principals engaged and tools and expectations for collaborative activity on using data for improvement, together with monitoring of effectiveness.	
		Gallimore et al (2009) Grade-level teams in experiment schools using an inquiry-focused protocol to solve instructional problems significantly increased achievement. Meetings of the teams used a protocol which provided guidance for an inquiry cycle of establishing a goal; planning; implementing; monitoring via common assessments; evaluating and moving to next cycle (USA).	
		Grek and Ozga (2012) key part of the inspectors' judgement in Scottish system is 'how well they know themselves' (through SSE)…knowledge production is seen as crucial and supporting schools to be learning organisations is a key aspect (Scotland, non-empirical).	

		Halverson et al (2005) Data was discussed in groups:e.g."In Pearson School …the principal and several lead teachers met regularly to develop reports on student learning collected through regular testing and anecdotal information. The discussions surrounding report generation and results provided an occasion to develop shared understanding of purpose and strong professional community among the leadership team." (p 21). Opportunities were planned into the organisation for staff to meet to reflect collaboratively in teams on the data and use it to plan improvement both at school level and also through 'district retreats' to formulate district goals. There was ongoing dialogue between teachers and leaders on the use and implications of data (USA).	
		Herman and Gribbons (2001) The researchers initially worked with school teams on data analysis, using district longitudinal data and any other data available, to produce a report which summarised 'how well are we doing?' 'are we well serving all subgroups?' As well as test and demographic data, there was also data about student attendance and survey data on indicators of school processes, such as teachers'use of time, parental involvement, extent to which teachers felt involved in decision making. The report describes in detail the steps taken in school group A, to use the initial report to identify underperforming sub- groups, initiate further data collection and to formulate a goal and strategy for improving learning, through adjustments to the mathematics curriculum and pedagogies. The processes for data use for school improvement became integrated into the school's way of working (USA).	
		Honig and Ikemoto (2008) studied collaborative data use in three districts, with one much more successful that the others. This was where there was a vision for school improvement through data use, tools and professional development, collective	

		identification of priorities and active engagement of principals (USA)	
		Kallemeyn (2014) In a schoolswith strengths in data use, organizational routines facilitated teachers' data use: collaborative teams and processes of inquiry. These routines stored knowledge about the types of data teachers ought to notice, and to a lesser extent, how they ought to interpret data and construct implications for practice (USA).	
		Lachat and Smith (2005) Leadership and support, including a time allocation, for collaborative working were found to be critical to success (USA)	
		Lai and McNaughton (2013) Data discussions in professional learning communities were an important component of research and development interventions. he importance of inter-dependence between schools and external experts, greater pedagogical content knowledge to link classroom instruction to achievement results and the creation and use of school artefacts (e.g., data analysis reports) to facilitate effective data use were found to be significant factors (New Zealand).	
		Lai and Hsiao (2014) About two thirds of school clusters could produce high quality self- evaluation data with support (New Zealand).	
		Marsh and Farrell (2015) In the schools studied practices employed were needs assessment, modelling, monitoring changes, use of dialogue and questioning, artifacts such as data repots, rules and expectations to support open, critical enquiry around data (USA)	
		Plowright (2007) SSE leads to the development of a learning organisation with 'individual and team learning …an integral part of organisational learning, potentially producing a school able to successfully respond to both internal and external changes' (p388)(England).	

		Schildkamp and Ehren (2013) provide the following structured approach for a data team: defining the problem, coming up with hypotheses concerning what causes the problem, collecting data to test the hypotheses, analyzing and interpreting data, drawing conclusions and implementing measures to improve..." (The Netherlands)	
		Schildkamp et al (2014) Where collaboration was common, as in England, Lithuania and Poland, interviewees were able to provide more concrete examples of how collaboration occurred and of the impact of decisions made based on the data discussed (England, Germany, Lithuania, the Netherlands, Poland)	
		Simons (2013) SSE needs to be built into the ongoing structure of the school with collaborative discussion and decision making (international, non-empirical).	
		Stringfield et al (2008) Contributory factors in schools using the high reliability principles and continuing to make academic progress are: sharing good practice at all levels through visiting and discussing others' lessons within and between schools; shared residential workshops and professional development to share good practice and barriers and discuss implications for future improvement. A contributory factor in schools continuing to make progress was succession planning to ensure that incoming headteachers were familiar with high reliability organisation principles (Wales).	
		Timperley and Parr (2009) Success was linked to needs analysis conducted with each school and drawing on a variety of evidence; goals for improvement were set by researchers and facilitators working with each school, use of professional learning communities to share data and implement data- informed change (New Zealand).	

		Wohlstetter et al (2008) Establishing a culture that valued and encouraged data use was a critical component of each system's efforts in obtaining shared objectives with the schools. School systems created explicit norms and expectations regarding data use at the system and school levels. Establishing a culture that valued and encouraged data use was a critical component of each system's efforts in obtaining shared objectives with the schools. School systems created explicit norms and expectations regarding data use at the system and school levels. Leaders at all levels co-constructed the vision and implementation of productive data-driven decision-making by creating an ethos of learning and continuous improvement rather than one of blame; leaders also distributed decision-making authority in a manner that empowered different staff members to utilise their expertise; the school systems directed their resources on building human and social capacity mainly by focusing on modelling and knowledge brokering amongst their staff; Accountability requirements for meeting standards and for school improvment plans that measure progress created strong incentives for schools to examine student achievement data. Systems examined compensation systems, and some tied them to student performance (USA) • Supporting teachers to use data appropriately and thoughtfully remained an ongoing challenge, with collaborative working with data considered essential, with built in collaboration time. Teachers relied on one another for support, new instructional strategies, and discussions about data. All of the school systems developed a discussion protocol to support collaborative working with data.	

		Young (2006) Role of leaders important in providing vision, expectations and norms for collaborative discussion of data to improve instruction across teams. Where this was not in place, although teachers may have used data to inform their own practice, teachers worked individually with little collaboration (USA).	