

Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment

Report written by Judy Sebba, Ruth Deakin Crick, Guoxing Yu,
Hilary Lawson, Wynne Harlen, Keren Durant

EPPI-Centre
Social Science Research Unit
Institute of Education
University of London

Report no. 1614 · October 2008

Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment

TECHNICAL REPORT

Review conducted working with the Review Group, practitioners,
EPPI-Centre members and international experts

Report by Professor Judy Sebba, University of Sussex
Dr Ruth Deakin Crick, University of Bristol
Dr Guoxing Yu, University of Bristol
Ms Hilary Lawson, University of Sussex
Professor Wynne Harlen, University of Bristol
Keren Durant, University of Bristol

The results of this systematic review are available in
four formats. See over page for details.

The results of this systematic review are available in four formats:

SUMMARY

Explains the purpose of the review and the main messages from the research evidence

REPORT

Describes the background and the findings of the review(s) but without full technical details of the methods used

TECHNICAL REPORT

Includes the background, main findings, and full technical details of the review

DATABASES

Access to codings describing each research study included in the review

These can be downloaded or accessed at
<http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2415&language=en-US>

The EPPI-Centre reference number for this report is 1614T.

This report should be cited as:

Sebba J, Crick RD, Yu, G, Lawson H, Harlen W, Durant K (2008) Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment. Technical report. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Authors of the systematic reviews on the EPPI-Centre website (<http://eppi.ioe.ac.uk/>) hold the copyright for the text of their reviews. The EPPI-Centre owns the copyright for all material on the website it has developed, including the contents of the databases, manuals, and keywording and data extraction systems. The centre and authors give permission for users of the site to display and print the contents of the site for their own non-commercial use, providing that the materials are not modified, copyright and other proprietary notices contained in the materials are retained, and the source of the material is cited clearly following the citation details provided. Otherwise users are not permitted to duplicate, reproduce, re-publish, distribute, or store material from this website without express written permission.

CONTENTS

Abstract	1
1. Background	5
<i>Aims and rationale for the review</i>	2
<i>Definitional and conceptual issues</i>	5
<i>Policy and practice background</i>	7
<i>The research background</i>	8
<i>Authors, funders and other users of the review</i>	8
2. Methods of the review	10
<i>Review methods</i>	10
<i>User involvement</i>	10
<i>Searching for, identifying and describing studies</i>	10
<i>Inclusion and exclusion criteria</i>	11
<i>In-depth review</i>	11
3. What research was found?	13
4. What were the findings of the studies	15
5. Implications	18
<i>Strengths and limitations of this review</i>	18
<i>Implications for policy</i>	18
<i>Implications for practice</i>	19
<i>Implications for research</i>	19
References	20
<i>Studies included in map and synthesis</i>	20
<i>Other references used in this report</i>	23
Appendices	25
<i>Appendix 1 Authorship of this report</i>	25
<i>Appendix 2 The standard EPPI-Centre systematic review process</i>	28

List of abbreviations

AAIA	Association for Achievement and Improvement through Assessment
AEI	Australian Educational Index
BEI	British Educational Index
DfES	Department for Education and Skills
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre
CRESST	Centre for Research on Evaluation, Standards and Student Testing
ERIC	Educational Resources Information Center
ICT	information and communication technology
IT	information technology
ITT	initial teacher training
MLA	Modern Languages Association
NFER	National Foundation for Educational Research
OFSTED	Office for Standards in Education
QCA	Qualifications and Curriculum Authority
REEL	Research Evidence in Education Library
RSA	Royal Society of Arts
SCRE	Scottish Council for Research in Education
SoSig	Social Science Information Gateway
TDA	Training and Development Agency for Schools
WoE	weight of evidence

Abstract

The review question

The overall question to be addressed in the review was:

What is the evidence of the impact on students in secondary schools of self and peer assessment?

In order to achieve all the aims of the review the further questions to be addressed were:

- How does any impact vary with the characteristics of the students and the approaches used in self and/or peer assessment?
- What conditions affect the impact of self and peer assessment?

What are the implications for assessment policy and practice of these findings?

Who wants to know and why?

Empirical research into student self and peer assessment has been concerned either with comparison of students' own assessment with teachers' assessment, or the effects of introducing self and peer assessment on students. This review is not concerned with the former but only with impact on students' academic achievement and non-cognitive outcomes. Several studies in an earlier review

by Black and Wiliam (Assessment Reform Group 1999, Black and Wiliam 1998a & b) reported gains in achievement of students who have been involved in self and peer assessment, but there is no existing systematic review of this field. The aim of the review was to fill this gap by addressing, through a systematic review, the research evidence of the impact on students in secondary schools of self and peer assessment. Evidence of how any impact depends on particular circumstances has been sought so that, where trustworthy evidence is found, implications for policy and practice can be identified.

Key agencies in the integrated children's services are expected to attend collaboratively to the well-being and growth of the learner as a person in a community. The Children Act (DFES 2004) and Children Plan (DCSF 2007) emphasise this and the five themes they espouse represent a range of factors and outcomes that should be attended to if learners are to take responsibility for themselves as lifelong learners. Putting the learner at centre stage in this process makes self and peer assessment a critical issue for both policy and practice because it builds upon students' self awareness, ownership of their own learning process and responsibility for their own learning. This review, focusing as it does on student self and peer assessment, will build on what is known by exploring evidence about the impact of this process on student outcomes.

Methods of the review

Ten electronic databases were searched and 19 key journals searched by hand. Review team members scanned reference lists, contacted key informants and organisations, and searched websites for research to include in the review. After examining the research in detail and assessing it for relevance and quality, the review's conclusions are based on an in-depth synthesis of 26 studies.

Key findings

Pupil outcomes

Most studies reported some positive outcomes for the following:

- Pupil attainment across a range of subject areas (9 out of 15 studies showed a positive effect)
- Pupil self-esteem (7 out of 9 studies showed a positive effect)
- Increased engagement with learning, especially goal setting, clarifying objectives, taking responsibility for learning, and/or increased confidence (17 out of 20 studies showed a positive effect)

Conditions that affect the impact of self or peer assessment

- The classroom culture was related to positive outcomes for students. The teacher needs to be committed to learners having control over the process, and to be able to discuss learning and develop effective student feedback.
- Self and peer assessment are more likely to impact on student outcomes when there is a move from a dependent to an interdependent relationship between teacher and students which enables teachers to adjust their teaching in response to student feedback.

- Although no clear relationship between students owning the process and positive outcomes was established in the review, it does seem to be important to involve students in 'co-designing' the criteria for evaluation. This helps them to develop a better grasp of their own strengths and weaknesses. Students need to be aware of the targets they are trying to achieve, and these should focus on outcome not process goals.
- There were no significant differences for different groups of students (for example by gender, ethnicity or prior attainment).
- There was no clear evidence to show whether peer and self assessment works better in some subjects than others, although limited evidence suggests that practice-based subjects may respond more immediately but that the outcomes are less embedded than in other subjects.

Strengths and limitations of this review

- The predominance of studies undertaken in the US (16) may limit the transferability of the findings to other countries. The variation between the assessment systems of different countries is likely to limit the potential for generalisation.
- The study design further limits the transferability of the findings of some of the studies. Just fewer than half the studies (11) involved control or comparison groups but five focused on only one class or group of students, suggesting the need for caution in generalising from these findings.
- A possible weakness in the studies reviewed relates to the very small number that sought consent from the participants in the research and the even smaller number that involved students in this. Only two studies (Bruce 2001, Goodrich 1997, both rated high on overall weight of evidence WoE) sought consent from students, one

also seeking consent from parents. A further three (Brookhart 2001, medium WoE; Crouch et al. 1997, medium WoE; Klein 1998, low WoE) sought consent from parents only. The other 21 studies either did not seek consent from anyone involved or did not report that they had done so. Establishing consent has become an increasing requirement within research ethics in recent years and many of these studies were published in the 1990s or earlier.

- A limitation of the review was the lack of involvement of students in the review process. Given the focus of the review, this might have been appropriate.
- A number of challenges were identified in undertaking effective self and peer assessment, and of evaluating it. It remained problematic to isolate the variables that contributed to any outcomes reported in order to demonstrate the effects specifically of self and peer assessment.

Implications for policy

The policy implications are concerned with ensuring greater emphasis on self and peer assessment within existing policies and making the relationships explicit rather than the creation of new or separate policies.

- The national primary and secondary strategies include coverage of personalised learning and assessment for learning that incorporate aspects of self and peer assessment. There is also discussion of group work in the materials that these strategies have made available to schools. It is clear from this review that students need to be taught both the skills of self assessment and those required to work with others if peer assessment is to be further developed. It appears that the dialogue involved in peer assessment in particular might be challenging but that peer assessment can help develop students' understanding of the requirements. In self assessment, no dialogue is involved with other students, but this understanding of requirements might

take longer since the student is pursuing this in isolation.

- Teachers need self and peer assessment issues to be further built into both initial training and continuing professional development. Increasingly, this emphasis will need to extend to the training and staff development of other staff involved in integrated children's services provision.
- The relationship between the outcomes of attainment and other outcomes such as 'enjoyment' and 'well-being' will need to be clearly articulated. The evaluation of these broader outcomes presents a challenge in terms of measurement.
- There was no evidence to support targeting of particular age, 'ability' or ethnic groups. The diverse range of pupils that these studies noted can benefit from self and peer assessment might suggest that such assessment can be a helpful context for enhancing inclusion. Sensitivity is needed to protect students from negative 'exposure' of any lack of progress or difficulties.

Implications for practice

- The review highlights the need for teacher commitment to learner control, developing a language for dialogue about learning and moving from a dependent to an interdependent relationship between teacher and students. Classrooms characterised by these processes will enable teachers to respond pedagogically to student feedback. This is at the heart of the personalising learning agenda.
- Seven studies identified the crucial need for students to receive some training in self assessment and to understand the terms and concepts which they are expected to use to assess themselves. While this has implications for building self and peer assessment into the national policies, it also suggests the need to build in these processes to day-to-day activities in classrooms.

- One study reported the influence of parents on pupils' own judgements of their work and identified the importance of parents being given a broader view of outcomes beyond grades. While this is derived from limited evidence, it suggests a need for more dialogue between parent, teacher and student.
- Pupils' understanding of progression and how this is enhanced through self and peer assessment
- Developing measures relating to the Every Child Matters (the national framework to support the joining up of children's services) outcomes and evaluating the impact of self and peer assessment longitudinally on these wider outcomes

Implications for future research

Future areas of research emerging from this review include the following:

- Detailed analysis of the cognitive mechanisms underlying self and peer assessment and the relationship between these and self regulation
- Comparisons of the development of self and peer assessment
- The impact of staff development in self and peer assessment for the school workforce

CHAPTER ONE

Background

Aims and rationale for the current review

Several studies in an earlier review by Black and Wiliam (Assessment Reform Group 1999, Black and Wiliam 1998a & b) reported gains in achievement of students who have been involved in self and peer assessment, but there was no existing systematic review of this field. The aim of the review was to fill this gap by conducting a systematic review which identified the research evidence on the impact of students' self and peer assessment in secondary schools. Evidence of the particular circumstances that influences impact was sought so that implications for policy and practice could be identified.

Arguments in favour of involving learners in the assessment of their own learning relate to theories of learning, the recognition of the importance of motivation for learning and the value of non-cognitive outcomes such as are needed to prepare students for lifelong learning (Deakin Crick et al, 2007).

Central to any notion of personalised learning or a learner-centred culture, is that the learners themselves should want to learn, and to become aware of themselves as learners, able to take responsibility for their own learning trajectories whether in or out of school and over a life span. Flutter and Rudduck (2004) argued, that in spite of decades of educational reform, students today might still agree with

Blishen (1969), that learning in school amounts to being 'told what to do and how to do it'. They say:

Although young peoples' lives have clearly changed in many ways, schooling continues to be based upon conceptions of childhood that regard young people as dependent and incapable. (Flutter and Rudduck 2004, p133)

Changes in the goals of education, needed to match the changes in society and to prepare future citizens for continued learning throughout life include for example, flexibility and new study skills. It follows that the more learners know about, and participate in, decisions about the goals of their own learning, the more they can direct their own learning efforts effectively. Self and peer assessment can enhance the achievement of the goals of education regarded as essential in preparing students to adapt to the accelerating changes in types of occupation and ways of living.

Definitional and conceptual issues

The goals of learning

Assessment in the context of education involves deciding, collecting and making judgements about evidence related to the goals of the learning being assessed. This review takes a broad view of the goals of learning, one that is reflected in the outcomes identified in the *Every Child Matters* outcomes:

Be healthy

Stay safe

Enjoy and achieve

Make a positive contribution

Achieve economic well-being

(DCSF 2007).

Student outcomes in the context of schooling include the knowledge, skills, understanding, values, attitudes and dispositions that are encompassed by the purposes of education and reflected in statutory frameworks. In this review we use the narrower term ‘attainment’ to refer to the attainment of a particular target set within the subjects of the curriculum, whereas we use the term ‘achievement’ to refer to a broader goal of education which includes both personal and social outcomes and attainment.

Formative and summative assessment

How the processes of assessment are conducted varies with the purpose of the assessment and those involved in carrying it out. The purpose may be summative, to assess the learning achieved at a particular time, or formative, to help on-going learning.

Self and peer assessment

Self assessment means that students make judgements about their own achievement and learning processes and take part in decisions about action for further progress in learning. In order to do this, they need to have a clear grasp of the goals of the learning and of the criteria to be applied in judging how well the goals have been attained. In self assessment, the distinction between formative and summative is often blurred since the feedback on performance is immediate; learners do not have to wait for someone else to tell them how well they have learned. But the extent to which it is used formatively will depend on the learners’ understanding of, and commitment to, the goals and on their ability to identify and take action necessary to take the next step in their learning.

Peer assessment involves students in assessing each other’s work, again through reflection on the goals and what it means to achieve them. It may take place in groups, where the aim may be as much the development of group processes as the promotion or judgement of individual learning. It may also take place in pairs. In the case of summative assessment, the learners reflect on and judge how well their performance meets certain criteria relating to the goals of the work.

The terms self-evaluation and peer-evaluation are used in some literature. These are interchangeable with self and peer assessment, the terms used in this review.

Types of impact on students

In this review we distinguish between three types of impact on students of the process of self and peer assessment. These are outcomes relating to attainment, outcomes relating to self esteem and outcomes relating to learning to learn. For example, the explicit goals of the self assessment might include identifying specific areas for future improvement. If this is achieved, it might be regarded as an important outcome relating to learning to learn. However, the maths, literacy or other subject-specific skills that were the focus of self assessment may have made little or no progress, suggesting limited outcomes relating to attainment.

The review question

The overall question to be addressed in the review was:

What is the evidence of the impact on students in secondary schools of self and peer assessment?

In order to achieve all the aims of the review the further questions to be addressed were:

- How does any impact vary with the characteristics of the students in secondary schools and the approaches used in self and/or peer assessment?

- What conditions affect the impact of self and peer assessment?
- What are the implications for assessment policy and practice of these findings?

Policy and practice background

A focus on learners and learning is now a central theme of policy and practice in education. In all four constituent parts of the UK, assessment for learning has become incorporated into the mainstream education policies. As discussed more fully in Daugherty and Ecclestone (2006), the contexts in each are distinct with respect to assessment more generally. In Wales, Scotland and Northern Ireland, the emphasis on testing has been reduced and assessment for learning is central in assessment policies. In Scotland, the *Assessment is for Learning* initiative has supported teachers to develop their classroom practice and informed central policy.

The concept of ‘personalised learning’ has emerged as a major focus for schools in England. The DCSF identifies it as an overarching idea with five key components (www.standards.dcsf.gov.uk/personalisedlearning/). These are:

- Assessment for learning
- Effective learning and teaching
- Curriculum entitlement and choice
- Organising the school
- Beyond the classroom

These are further expanded with illustrative examples in a number of publications drawing on research (e.g. Pollard and James 2004, Sebba et al. 2007).

In England, the government’s National Strategies, the National College for School Leadership, the Training and Development Agency for Schools (TDA) and the

Qualifications and Curriculum Authority (QCA) are moving these themes forward, but perhaps the most influential component to date has been the ‘assessment for learning’, or formative assessment which has a strong research pedigree (Black and Wiliam 1998a & b, Assessment Reform Group 1999, Assessment Reform Group 2002, Black et al. 2003). The aim of formative assessment is for the student to identify where they are in relation to the goals of learning and then to take the action necessary to work towards these goals. ‘In this view, self assessment is a sine qua non for effective learning’ (Black and Wiliam 1998a, p26).

Citizenship education also draws attention to learners themselves. There is evidence from a systematic review into the impact of citizenship education on the provision of schooling and on learning and achievement (Deakin Crick et al. 2004) that student choice and voice are key elements of pedagogy appropriate for citizenship education. Engaging with values, becoming helpfully involved in the community and becoming politically literate all foreground the learner as a person and their capacity to take responsibility for their own learning and development.

Attention to the learner as a person requires joined up thinking by key agencies in education and beyond. The Children Act (DFES 2004) is designed to do this and its five themes represent a range of factors and outcomes that should be attended to if learners are to take responsibility for themselves as lifelong learners. Putting the learner centre stage makes self and peer assessment a critical issue for both policy and practice because it builds upon student self awareness, student ownership and responsibility for their own learning. This review, focusing as it does on student self and peer assessment, builds on what is known by assessing the evidence about the impact of this process on student outcomes.

The research background

Empirical research into student self and peer assessment either has focused on comparing students' own assessment with that of teachers, or with the effects on students of introducing self and peer assessment. This review is not concerned with the former but only with the effects, both academic and non-academic, on secondary school students. Inevitably, there is considerable interest in whether levels of performance are raised by self and peer assessment.

McDonald and Boud (2003), in what is claimed to be a unique study of large scale introduction of self assessment across a range of subjects, reported positive changes associated with training in self assessment. In 10 high schools in the West Indies, teachers were trained in self assessment practices and introduced these to a group of students studying for external examinations. Their performance was compared with that of a matched control group of students, who were not given training in self assessment. The results showed a significant difference between the two groups, in favour of those trained in self assessment, with greatest effect sizes for business studies and the humanities and least for science subjects.

Black et al (2003) discussed how differences among subject disciplines may affect how teaching and learning take place and may account for the difference in impact of attempts to foster self-regulation through self and peer assessment. They reported larger effect sizes than McDonald and Boud but their intervention extended beyond self and peer assessment to other components of formative assessment such as questioning and sharing objectives.

Black and Wiliam's (1998a & b) review of classroom assessment included studies of the effect of training students with learning difficulties in self-monitoring. Students who received feedback through self-monitoring performed better than those who did not experience such feedback (Sawyer et al. 1992)

and those with self-monitoring did better than those with feedback only from the teacher (McCurdy and Shapiro 1992). Other studies found positive changes due to introducing self-scoring of tests (Masqud and Pillai 1991), and helping students to recognise how their self assessment differed from the judgments of others (Merrett and Merrett 1992).

Most studies reporting non-cognitive impact rely on self reporting but Schunk (1996) used a goal orientation inventory and a self-efficacy scale to demonstrate that learned self assessment was an overwhelming factor accounting for differences in mathematical skills achieved, beyond those noted from the manipulation of goal orientation.

Student self assessment and choice in learning are central themes that support the ecology of learning (Deakin Crick et al. 2007). A systematic review into citizenship provided evidence that student choice and participation in learning are key elements of pedagogy which support active citizenship (Deakin Crick et al. 2004, Deakin Crick et al. 2005). According to McCombs and Lauer (1997), when teachers derive their practices from a learner-centred perspective, they:

- include learners in decisions about how and what they learn and how that learning is assessed
- value each learner's unique perspectives
- respect and accommodate individual differences in learners' backgrounds, interests, abilities, and experiences
- treat learners as co-creators and partners in the teaching and learning process

Authors, funders and users

This review is the fifth EPPI-Centre review carried out by the Assessment and Learning Research Synthesis Group (ALRSG). Current members of the Review Group and overseas advisers are listed above. The review was based at the Graduate School of Education of the University of Bristol and the work is shared with the School of Education of the University of Sussex. The joint directors are Dr. Ruth Deakin Crick, at Bristol and Professor Judy Sebba, at Sussex. Professor Wynne Harlen acted as a consultant. The review was funded solely by the contract between the EPPI-Centre at the Institute of Education, University of London, and the University of Bristol, on behalf of the ALRSG. The review was carried out by the Review Team with the guidance of the ALRSG with the participation of its members, including teacher and adviser members, at various stages.

CHAPTER TWO

Methods of the review

Review Methods

The procedures used for searching, selecting, keywording, analysing and reporting were those of the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) and can be found in Appendix 1 and in the Technical Report of this review (www.eppi.ioe.ac.uk).

User involvement

The direct involvement of users in the conduct of the review was through their membership of the Review Group. This included a secondary school deputy head teacher with responsibility for assessment, a local authority primary adviser and a project director of the National College of School Leadership. Two members of the group were members of the Association for Achievement and Improvement through Assessment (AAIA), another led the review of assessment in Wales and another was Director of the Learning to Learn project of the ESRC's Teaching and Learning Research Programme. Eight of the Review Group were members of the Assessment Reform Group. The Review Group had regular contact with the DFES who funded this review and with the Qualifications and Curriculum Authority (QCA).

Users were involved in developing the protocol, identifying review specific keywords, reviewing the map of the research and identifying

exclusion and inclusion criteria for selecting studies for in-depth review. They provided information about studies through personal contact and provided verification of possible interpretations of the emerging findings. One user member of the review group keyworded two studies.

Searching for, identifying and describing studies

Studies were identified from bibliographic databases, and through searches on full texts of journal publishers' web pages including both current and archived journals, handsearching of key journals in education, citation searches of key authors, reference lists of key authors and papers, references on key specialist websites such as NFER, CRESST and SCORE and personal contacts and direct requests to key researchers in the area of self and peer assessment.

Searches were limited to studies conducted in a specific time period of 1980-2005. This starting date was selected so that studies in the early 1980s of Records of Achievement and Profiling, developed at that time, could also be included. Bibliographic databases and journals primarily in languages other than English were not searched. Terms were used 'free text' in the search and all searches were recorded.

Pairs of review group members working independently applied the inclusion and

exclusion criteria and completed keywording and then compared their decisions. The EPPI-Centre carried out a quality-assurance role in both applying inclusion and exclusion criteria (20 of 214) and in keywording a sample of studies (18 of 51 studies). All studies for which full texts were obtained by November 2005 were keyworded, using both the generic and review-specific keywords and added to the larger EPPI-Centre database, REEL, for others to access via the website. The agreed keywords were used to produce the systematic map of the 51 included studies.

Inclusion and exclusion criteria

Searching and selection of studies was guided by the following inclusion criteria:

- *Language of the report:* Studies included were written in English.
- *Types of assessment:* Studies were included which dealt with the impact of some form of formative or summative assessment that involved students assessing their own work or that of their peers.
- *Context of assessment:* Studies were included from all curricular areas and related to the full range of learning processes including acquisition of skills and values and metacognition.
- *Study population and setting:* Initially, studies were included which dealt with self and peer assessment procedures used by students, aged 4-19, in school. For the in-depth review, this was limited to secondary schools only.
- *Study type and study design:* Studies were included if they reported quantitative or qualitative evidence of changes in students that could be ascribed to the self or peer assessment for formative or summative purposes. All study types were included if the focus was considered to be relevant.

Studies meeting some of the above inclusion criteria were **excluded** for the following reasons and labelled accordingly:

- A: Not self or peer assessment (excluded if students had no part in collecting and interpreting information about their performance).
- B: Not related to education in school (excluded if studies were related to college students; higher education; nursing education, other vocational) and for the in-depth review, secondary school.
- C: Not reporting impact on students of the process of self or peer assessment but just the outcome of the assessment itself.
- D: Not research (excluded if not empirical study of particular procedures of assessment by teachers; also excluded if only procedure development were reported or description without report of use; excluded if handbooks, textbooks and reviews). These were used to inform background context, but were not included in data extraction.

In-depth review

In consultation with the funders, it was agreed that the priority for this review should be the 26 studies related to secondary education. These were entered into EPPI-Reviewer and data-extracted by two people independently, using EPPI-Centre generic and review-specific questions relating to the weight of evidence of each study. For each study, those completing independent data extractions compared their decisions and came to a consensus by direct communication. Eight studies were data-extracted by a member of EPPI-Centre staff for quality-assurance purposes. Judgments were made using the EPPI-Centre 'weight of evidence' criteria in order to ensure that conclusions were based on the most sound and relevant evidence. This included three aspects of each study (A, B, C) and the combination of these to give an overall judgement of the weight (D) that could be attached to the evidence from a particular study to answer the review question.

The criteria for assessing weight were as follows:

A: Soundness of methodology

Judgement of how well the study had been carried out was informed by the responses to questions about the internal methodological coherence during the data extraction.

B: Appropriateness of research design and analysis used for answering the review questions

Judgement was made in relation to the extent to which the type and design of study enabled it to be used to address the specific review questions. This was not a judgement of the value of the study in its own right, but only in respect of how well its design enabled the review questions to be answered and was thus review-specific.

C: Relevance of the particular topic focus of the study for answering the review questions

Judgement was made about the match of the study to the purposes of the review and was not a judgement on the value of the study *per se*. In this case, the aspect of interest was how well the focus of the data collected helped to answer the review questions.

D: Overall weight taking into account A, B, and C

The judgements for the three aspects were combined into an overall weight of evidence towards answering the review questions. In doing this, where there was a difference of judgement between A, B and C, the overall judgement was based on the majority rating but with the condition that the overall weight could not be higher than the weight for C.

In all four aspects, the judgement is dependent on the quality of reporting in the study in that lack of information about research design or methods in the report does not necessarily mean these were poor in the study itself. In all four aspects studies were rated high, medium or low.

The weight of evidence assessments and the responses to the generic and specific review questions in the data extraction were used as a basis for producing a narrative synthesis to address the review questions.

CHAPTER THREE

What research was found?

The total number of papers screened was 214, and 51 of those were keyworded. The studies were mainly from the US with English-speaking countries totalling 45 of the 51 studies keyworded. Nearly all were interventions involving researcher manipulation. Nearly half of these explored relationships. There were a relatively small number of descriptive studies in this review compared to many reviews in education.

Half of the studies keyworded focused on formative assessment and all were characterised by one or more forms of self and/or peer assessment (as expected given inclusion criteria). Nearly three-quarters of the studies reported on an aspect of attainment with nearly half reporting on engagement in learning as an outcome, in addition, or instead of, attainment. A smaller number of studies reported 'social' aspects of learning such as enjoyment, confidence to participate and social engagement. The two outcomes reported in the least number of studies are 'well-being' as part of the *Every Child Matters* agenda and students' understanding of progression, that has emerged more recently as an issue in relation to progress in the English national curriculum.

One fifth of the studies keyworded did not have a subject specific focus, in most cases taking a cross-curricular approach to self and/or peer assessment. Of those that were subject-specific, the two most frequently covered subjects were English and mathematics, with other subjects reported much less often.

The studies were keyworded in terms of three levels of ownership: low, where there was no genuine ownership by the students of the assessment process, medium, where students adopt the goals and criteria identified by the teacher through consultation and high, where students determine the goals and are committed to engaging in learning to achieve them. The studies keyworded were evenly distributed across all three levels of ownership, 19 showing low levels of ownership, 14 medium and 18 high. In the 51 studies keyworded, higher levels of student ownership were found where both self and peer assessment approaches were used. This lends support to the view that self and peer assessment have an important contribution to make to the process of learning though what is shown here is an association, so no causal relationship can be assumed.

Summary of systematic map

From the 51 studies in the systematic map the following summary can be drawn:

- Studies were undertaken in a range of countries but with the majority from the US.
- Most studies were researcher-manipulated evaluations with just under half exploring relationships - for example, between different types of self and/or peer assessment and outcomes.

- Most studies focused on English or mathematics or had no subject-specific focus - for example, taking a cross-curricular approach.
- Most studies reported on attainment outcomes with nearly half reporting on engagement in learning, but social aspects of learning were relatively infrequently covered.
- While the levels of ownership by students of the assessment process varied across the studies, higher levels were noted by the reviewers for the studies in which both self and peer assessment featured. No causal relationship between these factors can be assumed.

CHAPTER FOUR

What were the findings of the studies?

Further details of studies included in the in-depth review

The outcomes of the searching and keywording confirmed that there was no shortage of studies in this area. More than half (16) of the 26 studies in the in-depth review were conducted in the US and nearly all the studies were researcher-manipulated evaluations. A wide range of study designs was employed. Nine studies made comparisons between intervention and control groups, four of which (Kitsantas et al. 2004, Klein 1998, Masqud and Pillai 1991, Ross et al. 1998) involved random allocation and a further two studies (Brookhart 2001, Deakin et al. unpublished), had comparison groups. Five studies (Gregait et al. 1997, Marshall 1993, Powell and Makin 1994, Uselman 1996, Young et al. 1997) only focused on one class or group of students.

The overall weight of evidence was high for eight studies, medium for eleven studies and low for seven studies. The studies that were rated low overall are included in the synthesis as they have specific contributions to make to addressing the research questions (e.g. Carter 1997 makes a contribution in looking specifically at students designated as gifted), but they are given less weight in the conclusions drawn. There was no relationship between use of control or comparison groups and overall weighting of evidence as low, medium or high but three of the five small-scale studies were assessed as low on overall weight of evidence.

Synthesis of evidence

The main characteristics of the 26 studies and outcomes reported are summarised in Table 4.1.

Three types of outcomes for pupils were identified from the synthesis: increased pupil attainment, improvements in self esteem and increased engagement with learning, often referred to in the literature as aspects of 'learning to learn' or more recently 'learning how to learn' (James et al. 2006).

Increased pupil attainment

Nine of the fifteen studies that reported performance outcomes noted increases in pupil attainment, though in one (Knubb-Manninen 1994, low WoE) this impact was weak. Four of the nine studies (Knubb-Manninen 1994, Powell and Makin 1994, Rief 1990, Uselman 1996) were rated low on overall weight of evidence. A tenth study (Kitsantas et al. 2004, medium WoE) reported an increase in attainment only when the self evaluation included outcome goals. Example of the types of outcomes reported included language scores, science tests and volley-ball skills.

Five studies (four rated high, one medium WoE) reported no significant increase in attainment or performance. For example, Katstra et al. (1987, high WoE) found that writing skills measured through word counts and

attitude scales were no better following self evaluation. Two studies noted that evaluations improved performance whether done by the teacher or the student. Olin and Sullivan (2002, high WoE) noted highest grades on work at post-test for students evaluated by the teacher only.

Improvements in self esteem

In keeping with previous research (e.g. Marsh et al. 2005), self esteem was accepted in some studies, as a proxy for subsequent attainment, acknowledging that there are wide variations in both definitions and robustness of the measures of self esteem. Five of the nine studies reporting outcomes on self esteem measured this through student completed questionnaires, while three used self reporting by students through interviews. Seven of these nine studies (one low, five medium, one high WoE) reported positive outcomes.

Improvements in learning to learn

Twenty of the 26 studies reported on outcomes in learning to learn. This area had both the most studies, and the most studies reporting positive findings. Furthermore, it is easier to report shorter term effects on learning skills than on attainment. Much of the data reported are based on perceptions of students and teachers. Seventeen (five high, eight medium, four low WoE) of the 20 studies reported positive outcomes on goal setting, clarification of objectives, increased responsibility for learning and/or increased confidence. For example, Brookhart (2001, medium WoE) noted that students used both summative and formative self assessment in their approach to study, transfer of learning and in self monitoring to positive effect.

Four studies reported improved study skills in particular, relating to setting goals and clarifying expectations and many studies reported that dialogue about learning increased. Peer assessment was noted by Klenowski (1995, medium WoE) to enhance self evaluation and interaction, by Katstra et al. (1987, high WoE) and Uselman (1996,

low WoE) to improve attitudes to learning, and by Powell and Makin (1994, low WoE) and Ross et al. (2002, medium Woe) to improve discussion and explanations about mistakes. Deakin Crick et al. (unpublished, medium WoE) found that while learning power increased following intervention, significant differences in outcomes emerged both between schools and between classes in the same school. Three studies concluded that learning to learn and/or metacognition did not increase following self assessment.

How do self and peer assessment work?

Explanations for how self and peer assessment might impact on pupil outcomes include Bruce's (2001, high WoE) suggestion that expectations of learning are clarified through the student co-designing the evaluation criteria with the teacher. This is confirmed by Klenowski's (1995, medium WoE) study in which the importance of teacher and student defining the criteria together is emphasised. Other factors that emerge include students developing a better grasp of their own strengths and weaknesses, becoming more accountable for their own learning, self and peer assessment providing a broader evaluation than test scores alone and improving the student-teacher dialogue, which in turn raises achievement.

In subsequent research which addresses the assessment of learning dispositions, teachers considered these to be a necessary prerequisite for attainment and a vehicle for taking responsibility for their own learning trajectories (Deakin Crick 2007). By engaging in self assessment, students were reported to be developing a language with which to describe the processes of learning, and this was an important foundation for self awareness and ownership.

Table 4.1 Which type(s) of study does this report describe? (not mutually exclusive, N=26)

Attribute	Number
Description	2
Exploration of relationships	9
Evaluation: naturally occurring	3
Evaluation: researcher-manipulated	22

Summary of the results of the synthesis

Student outcomes

Most studies reported some positive outcomes including for the following:

- Pupil attainment across a range of subject areas (9 out of 15 studies showed a positive effect)
- Pupil self esteem (7 out of 9 studies showed a positive effect)
- Increased engagement with learning, especially goal setting, clarifying objectives, taking responsibility for learning, and/or increased confidence (17 out of 20 studies showed a positive effect)

Conditions that affect the impact of self- or peer assessment

- The classroom culture was related to positive outcomes for students. The teacher needs to be committed to learners having control over the process, and to be able to discuss learning and develop effective student feedback.
- Self and peer assessment are more likely to impact on student outcomes when there is a move from a dependent to an interdependent relationship between teacher and students which enables teachers to adjust their teaching in response to student feedback.
- Although no clear relationship between students owning the process and positive

outcomes was established in the review, it does seem to be important to involve students in ‘co-designing’ the criteria for evaluation. This helps them to develop a better grasp of their own strengths and weaknesses. Students need to be aware of the targets they are trying to achieve, and these should focus on outcome not process goals.

- There were no significant differences for different groups of students (for example by gender, ethnicity or prior attainment).
- There was no clear evidence to show whether peer and self assessment works better in some subjects than others, although limited evidence suggests that practice-based subjects may respond more immediately but that the outcomes are less embedded than in other subjects.

Differences between self and peer assessment

- No clear differences emerged between the effects of self and peer assessment. Self assessment is sometimes assumed to be easier than peer assessment as it makes less demands on dialogue between students. Some studies suggested that self assessment is more demanding, as in peer assessment students can learn the skills of what and how to assess from one another.

CHAPTER FIVE

Implications

Strengths and limitations of this review

Potential limitations arising from the nature of the evidence include the type of studies and predominance of those undertaken in the US (16 studies) which may limit the transferability of the findings to other countries. The variation between the countries in particular, in the context of assessment systems, is likely to limit the potential for generalisation. Just less than half the studies (11) involved control or comparison groups but five focused on only one class or group of students, suggesting the need for caution in generalising from these findings.

A possible weakness in the studies reviewed relates to the very small number that sought consent from the participants in the research and the even smaller number that involved students in this. Only two studies (Bruce 2001, Goodrich 1997, both high WoE) sought consent from students, one also seeking consent from parents. A further three (Brookhart 2001, medium WoE; Crouch et al. 1997, medium WoE; Klein 1998, low WoE) sought consent from parents only. The other 21 studies either did not seek consent from anyone involved or did not report that they had done so. Establishing consent has become an increasing requirement within the wider frameworks on research ethics in recent years. Lack of coverage of consent may relate to the fact that the majority of studies in this review were published in the 1990s (two date back to the 1980s) but overall,

given the focus of the research on pupil involvement, this is a limitation.

A number of difficulties emerged from the studies reviewed both of undertaking effective self and peer assessment, and of evaluating it. It remained problematic to isolate specifically, the effects of self and peer assessment. Student capacity to evaluate themselves honestly was raised by the students as an issue in Bickmore's study (1981, medium WoE). One constraint raised by teachers was the time needed for reading and assessing student comments (Bickmore 1981, Carter 1997).

A limitation of the review itself was the lack of involvement of students in the review process. Given the focus of the review, this might have been appropriate. Other limitations include the possibility that studies were missed in the searching process and the obtaining of eight full texts after the cut-off date for keywording. Furthermore, the process of completing the review and publishing it means that studies published after November 2005 would not have been included in the review. While updating of the review can be undertaken in the future, it is possible that relevant studies published in the intervening period were not included.

Implications for policy

The policy implications are less about the creation of new or separate policies and more

about ensuring greater emphasis on self and peer assessment within existing policies and making the relationships with existing policies more explicit.

- The national primary and secondary strategies include coverage of personalised learning and assessment for learning that incorporates aspects of self and peer assessment. It is clear from this review that students need to be taught both the skills of self assessment and those required to work with others if peer assessment is to be further developed. It appears that the dialogue involved in peer assessment in particular might be challenging, but that peer assessment can help develop students' understanding of the requirements.
- Teachers need pupil self and peer assessment issues to be further built into both initial training and continuing professional development. Increasingly, this emphasis will need to extend to the training and staff development of other staff involved in integrated children's services provision.
- The relationship between the outcomes of attainment and other outcomes such as 'enjoyment' and 'well-being' will need to be clearly articulated if students are to self and peer assess these.
- There was no evidence to support targeting of particular age, 'ability' or ethnic groups. The diverse range of pupils that these studies noted can benefit from self and peer assessment might suggest that such assessment can be a helpful context for enhancing inclusion. Sensitivity is needed to protect students from negative 'exposure' of any lack of progress or difficulties.

Implications for practice

- The review highlights the need for teacher commitment to learner control, developing a language for dialogue about learning and moving from a dependent to an interdependent relationship between teacher and students. Classrooms characterised

by these processes will enable teachers to review their teaching in the light of student feedback. This is at the heart of the personalising learning agenda.

- Seven studies identified the crucial need for students to receive some training in self assessment and to understand the terms and concepts which they are expected to use to assess themselves. While this has implications for building self and peer assessment into the national policies, it also suggests the need to build these processes into day-to-day activities in classrooms.
- One study reported the influence of parents on pupils' own judgements of their work and identified the importance of parents being given a broader view of outcomes beyond grades. While derived from limited evidence, it is worth considering the opportunities for dialogue between parent, teacher and student.

Implications for research

Future areas of research emerging from this review include the following:

- Detailed analysis of the cognitive mechanisms underlying self and peer assessment and the relationship between these and self regulation
- Comparisons of the development of self and peer assessment
- Pupils' understanding of progression and how this is enhanced through self and peer assessment
- Developing measures relating to the Every Child Matters outcomes and evaluating the impact of self and peer assessment longitudinally on these wider outcomes
- The impact of staff development in self and peer assessment for the school workforce

References

Studies included in map and synthesis

Those included within the in-depth review are asterisked

Ackerman S, Hughs L, Wilder R (1997) *Improving student responsibility. Master's Action Research Project, Saint Xavier University and IRI/Skylight*. Available at: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/29/ba/54.pdf

Adams N, Cooper G, Johnson L, Wojtysiak K (1996) *Improving student engagement in learning activities* (Report No. PS 024590). Lincolnshire, Illinois, USA: Saint Xavier University. (ERIC Document Reproduction Service No. ED 400 076)

* Bickmore DK (1981) The effects of student self-evaluation and pupil-teacher conferences on student perceptions, self concepts and learning. *Dissertation Abstracts International* 42: 1927.

Bielecki DR (1988) Reader-based and criterion-based peer evaluation in the writing performance and revising practices of twelfth graders. *Dissertation Abstracts International* 48: 2262A.

* Brookhart SM (2001) Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy and Practice* 8: 153-169.

Brookhart SM, Andolina M, Zuza M, Furman R (2004) Minute math: an action research study of student self assessment. *Educational Studies In Mathematics* 57: 213-227.

* Bruce LB (2001) Student self assessment: encouraging active engagement in learning. *Dissertation Abstracts International Section A: Humanities and Social Sciences* 62: 1309.

* Carter CR (1997) Assessment: shifting the responsibility. *Journal of Secondary Gifted Education* 9: 68-75.

Connell MC, Randall C, Wilson J, Lutz S, Lamb D R(1993). Building independence during in-class transitions: teaching in-class transition skills to preschoolers with developmental delays through choral-response-based self assessment and contingent praise. *Education and Treatment of Children* 16: 160-174.

* Crouch DW, Ward P, Patrick CA (1997) The effects of peer-mediated accountability on task accomplishment during volleyball drills in elementary physical education. *Journal of Teaching in Physical Education* 17: 26-39.

Cunningham J, Krull C, Land N, Russell S (2000) Motivating students to be self-reflective learners through goal-setting and self-evaluation Master of Arts action research project. Saint Xavier University and Skylight Professional Development. Downloaded from: <http://eric.ed.gov/>

ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/16/9c/4b.pdf

* Deakin Crick R, Broadfoot P, Claxton G, Yu G (unpublished) Developing 'learning power' in the classroom: the Effective Lifelong Learning Inventory (ELLI) in practice. Bristol: University of Bristol, Graduate School of Education.

* Ezell D, Klein CE, Ezell-Powell S (1999) Empowering students with mental retardation through portfolio assessment: a tool for fostering self-determination skills. *Education And Training In Mental Retardation And Developmental Disabilities* 34: 453-463.

Fernandes M, Fontana D (1996) Changes in control beliefs in Portuguese primary school pupils as a consequence of the employment of self assessment strategies. *British Journal of Educational Psychology* 66: 301-313.

Fontana D, Fernandes M (1994) Improvements in mathematics performance as a consequence of self assessment in Portuguese primary school pupils. *British Journal of Educational Psychology* 64: 407-417.

Freed-Garrod J (1999) Assessment in the arts: elementary-aged students as qualitative assessors of their own and peers' musical compositions. *Bulletin- Council For Research In Music Education* 50-63.

Fuchs LS, Fuchs D, Prentice K, Burch M, Hamlett CL, Owen R, Schroeter K (2003) Enhancing third-grade student' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology* 95: 306-315.

Gerhart GL (1986) Effects of evaluative statements on artistic performance and motivation. *Studies in Art Education* 27: 61-72.

* Goodrich HW (1997) Student self assessment: at the intersection of metacognition and authentic assessment. *Dissertation Abstracts International Section A: Humanities and Social Sciences* 57: 2854.

* Gregait LH, Johnsen DR, Nielsen PS (1997) Improving evaluation of student participation in physical education through self assessment. Xavier University and IRI/Skylight. Downloaded from: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/15/1f/a6.pdf

Hallahan DP, Lloyd JW, Kneedler RD, Marshall KJ (1982) A comparison of the effects of self-versus teacher-assessment of on-task behavior. *Behavior Therapy* 13: 715-723.

* Hewitt MP (2002) Self-evaluation tendencies of junior high instrumentalists. *Journal of Research in Music Education* 50: 215-226.

Johnson DW, Johnson RT, Stanne MB, Garibaldi A (1990) Impact of group processing on achievement in cooperative groups. *Journal of Social Psychology* 130: 507-516.

* Katstra J, Tollefson N, Gilbert E (1987) The effects of peer evaluation on attitude toward writing and writing fluency of ninth grade students. *Journal of Educational Research* 80: 168-172.

* Kitsantas A, Reiser RA, Doster J (2004) Developing self-regulated learners: goal setting, self-evaluation, and organizational signals during acquisition of procedural skills. *Journal of Experimental Education* 72: 269-287.

* Klein DCD (1998) *I've seen this before? The effects of self-monitoring and multiple context instruction on knowledge representation and transfer among middle school students.* CSE Technical Report 466. Los Angeles: National Center for Research on Evaluation, Standards and Student Testing (CRESST), University of California, Los Angeles.

* Klenowski V (1995) Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education* 2: 145-163.

* Knubb-Manninen G (1994) Processing habits and second language learning: students' self-evaluation of their learning. *Language Culture And Curriculum* 7: 145-159.

Lee D, Gavine D (2003) Goal-setting and self assessment in year 7 students. *Educational Research* 45: 49-59.

* Marshall J (1993) Self assessment in secondary schools: a study. In: *The ACSA '93 collection conference papers: curriculum in profile: quality or inequality?* Volume 2, pages 843-864. Belconnen ACT: Australian Curriculum Studies Association (ACSA). Brisbane: Queensland University of Technology

* Masqud M, Pillai CM (1991) Effect of self-scoring on subsequent performances in academic achievement tests. *Educational Research* 33: 151-154.

McCurdy BL, Shapiro ES (1992) A comparison of teacher-, peer-, and self-monitoring with curriculum-based measurement in reading among students with learning disabilities. *Journal of Special Education* 26: 162-180.

* McDonald B, Boud D (2003) The impact of self assessment on achievement: the effects of self- assessment training on performance in external examinations. *Assessment in Education* 10: 209-220.

Nelson JR, Smith DJ, Colvin G (1995) The effects of a peer-mediated self-evaluation procedure on the recess behavior of students with behavior problems. *Remedial And Special Education* 16: 117-124.

* Olina Z, Sullivan HJ (2002) Effects of classroom evaluation strategies on student achievement and attitudes. *Educational Technology Research and Development* 50: 61-75.

* Olina Z, Sullivan HJ (2004) Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development* 52: 5-22.

Pigott H, Fantuzzo JW, Heggie DL, Clement PW (1984) A student-administered group-oriented contingency intervention: Its efficacy in a regular classroom. *Child and Family Behavior Therapy* 6: 41-55.

* Powell SD, Makin M (1994) Enabling pupils with learning difficulties to reflect on their own thinking. *British Educational Research Journal* 20: 579-593.

* Rief L (1990) Finding the value in evaluation: self assessment in a middle school classroom. *Educational Leadership* 47: 24-29.

Ringle JS (1987) Self-evaluation: a learned self-regulatory process. *Dissertation Abstracts International* 47: 3374.

Ross JA, Hogaboam Gray A, Rolheiser C (2002) Student self-evaluation in Grade 5-6 mathematics effects on problem-solving achievement. *Educational Assessment* 8: 43-59.

* Ross JA, Rolheiser C, Hoaboam-Gray A (1998) Impact of self-evaluation training on mathematics achievement in a cooperative learning environment. Paper presented at the Annual Meeting of the American Educational Research Association San Diego, April 1998. Downloaded from: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/15/b5/eb.pdf

Ross JA, Rolheiser C, Hogaboam-Gray A (1999) Effects of self-evaluation training on narrative writing. *Assessing Writing* 6: 107-132.

* Ross JA, Rolheiser C, Hogaboam-Gray A (2002) Influences on student cognitions about evaluation. *Assessment in Education* 9: 81-95.

Schunk DH (1995) Self-monitoring of skill acquisition through self-evaluation of capabilities. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 1995. Downloaded from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/13/cc/e8.pdf

Schunk DH (1996) Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal* 33: 359-382.

Shaw AH (1997) Improving literacy for first graders through writing for authentic purposes and self assessment. EdD Practicum, Nova Southeastern University. Downloaded from: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/ec/51.pdf

* Uselman J (1996) Aiding seventh and eighth graders at a private Christian school to take ownership of their own learning. EdD Practicum Report, Nova Southeastern University. Downloaded from: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/c3/e4.pdf

van Kraayenoord CE, Paris SG (1997) Australian students' self-appraisal of their work samples and academic progress. *Elementary School Journal* 97: 523-537.

Young EA (2000) Enhancing student writing by teaching self assessment strategies that incorporate the criteria of good writing. *Dissertation Abstracts International, Section A: The Humanities and Social Sciences* 61: 1299.

* Young JP, Mathews SR, Kietzmann AM, Westerfield T (1997) Getting disenchanted adolescents to participate in school literacy activities: portfolio conferences. *Journal of Adolescent and Adult Literacy* 40: 348-60.

Other references used in this report

Assessment Reform Group (1999) *Assessment for learning: beyond the black box*. Cambridge: University of Cambridge School of Education.

Assessment Reform Group (2002) *Testing, motivation and learning*. Cambridge: Assessment Reform Group.

Black P, Harrison C, Lee C, Marshall B, Wiliam, D (2003) *Assessment for Learning: putting it into practice*. Maidenhead, Open University Press.

Black P, Wiliam D (1998a) Assessment and classroom learning. *Assessment in Education* 5: 7-74.

Black P, Wiliam D (1998b) *Inside the black box*. London: King's College London School of Education.

Blishen E (1969) *The school that I'd like*. Harmondsworth: Penguin.

DCSF (2007) *The children's plan: building brighter futures*. Norwich: HMSO.

Daugherty R, Ecclestone K (2006) Constructing assessment for learning in the UK policy context. In: J.Gardner (ed.) *Assessment and learning*. London: Sage

Deakin Crick R, McCombs B, Hadden A, Broadfoot P, Tew M (2007) The ecology of learning: factors contributing to learner centred classroom cultures. *Research Papers in Education* 22: 267-307.

Deakin Crick R, Coates M, Taylor M, Ritchie S (2004) *A systematic review of the impact of citizenship education on the provision of schooling*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Deakin Crick R, Tew M, Taylor M, Durant K, Samuel E (2005) *A systematic review of the impact of citizenship education on learning and achievement*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Deakin Crick R (2007) Learning how to learn: the dynamic assessment of learning power. *Curriculum Journal* 18: 135-153.

DFES (2004) *The Children Act*. London: HMSO.

DFES (2005) *Every child matters*. London: DFES. Downloaded from: www.everychildmatters.gov.uk

Flutter J, Rudduck J (2004) *Consulting pupils: what's in it for schools?* London: Routledge Falmer.

Marsh HW, Trautwein U, Lüdtke O, Köller O, Baumert J (2005) Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Development* **76**: 397-416.

McCombs B, Lauer P (1997) Development and validation of the learner-centred battery: self assessment tools for teacher reflection and professional development. *The Professional Educator* **20**: 1-20.

Merrett J, Merrett F (1992) Classroom management for project work: an application of correspondence training. *Educational Studies* **18**: 3-10.

Pollard A, James M (eds) (2004) *Personalised learning: a commentary by the teaching and learning research programme*. Cambridge: ESRC TLRP.

Sawyer R, Graham S, Harris K (1992) Direct teaching, strategy instruction and strategy instruction with explicit self-regulation: effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology* **84**: 340-350.

Sebba J, Brown N, Stewart S, Galton M, James M (2007) *An investigation of personalised learning approaches used by schools*. Nottingham: DfES Publications.

Appendix 1.1: Authorship of this report

This work is a report of a systematic review conducted by the Assessment and Learning Research Synthesis Group (ALRSG).

The authors of this report, and members of the Review Group, are:

Professor Judy Sebba, University of Sussex
Dr Ruth Deakin Crick, University of Bristol
Dr Guoxing Yu, University of Bristol
Ms Hilary Lawson, University of Sussex
Professor Wynne Harlen, University of Bristol
Keren Durant, University of Bristol

For further information about this review, please contact:

Judy Sebba
Sussex Institute
University of Sussex
Falmer BN1 9QQ

Tel: +44 01273 678260
Email: j.c.sebba@sussex.ac.uk

For further information about the work of the EPPI-Centre, please contact:

EPPI-Centre
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square

Tel: +44 (0)20 7612 6397
Fax: +44 (0)20 7612 6800
Email: EPPIAdmin@ioe.ac.uk

Review group

ALRSG members

Professor Paul Black, King's College London
Professor Patricia Broadfoot, University of Gloucestershire
Professor Richard Daugherty, Aberystwyth University
Professor Kathryn Ecclestone, Oxford Brookes University
Professor John Gardner, Queen's University, Belfast
Professor Wynne Harlen, University of Bristol
Professor Mary James, University of Cambridge
Dr Gordon Stobart, Institute of Education, University of London

(The above are also members of the Assessment Reform Group.)

Practitioners (affiliations at time of review)

Mr P. Dudley, Head of School Improvement and Lifelong Learning, Redbridge, and member of AAIA
Mr R. Bevan, Deputy Headteacher, King Edward VI Grammar School, Chelmsford
Ms P. Rayner, Link Adviser for Primary Education, Nottinghamshire.

EPPI-Centre members

Professor David Gough, Director
Ms Zoe Garrett
Ms Kelly Dickson
Dr Mark Newman

International experts (affiliations at time of review)

Dr Steven Bakker, ETS International, the Netherlands
Dr Dennis Bartels, President, TERC, Cambridge, Massachusetts, USA
Professor Lorrie Shepard, President, AERA, 1999-2000, University of Colorado
Professor Eva Baker, co-director of CRESST, University of California, Los Angeles, USA
Dr T. Crooks, Director, EARM, University of Otago, Dunedin, New Zealand
Professor Dylan Wiliam, Educational Testing Service, USA

Institutional base

Centre for Equity in Education
School of Education
A5.17 Ellen Wilkinson Building
University of Manchester
Oxford Road
Manchester
M13 9PL

Conflict of interest

A potential conflict of interest in this review is the fact that two of the authors of the review are also authors of one of the studies in it. The steps taken to minimise this risk were:

- the identification of papers through searching and keywording by the Review Group, with this

paper being identified and keyworded by other members of the team;

- a declaration of interest to the Review Group;
- the double data extraction and moderation by other members of the Review Group.

Acknowledgements

This review was carried out with funding from the Department for Education and Skills (DfES) through the EPPI-Centre. We would also like to acknowledge the time and support provided by the institutions with which the review team members are associated. The help and advice of the EPPI-Centre staff who were linked to this review, Kelly Dickson, Mark Newman, Zoe Garrett and David Gough, is gratefully acknowledged. Thanks to the two anonymous reviewers for their constructive suggestions.

Appendix 2: The standard EPPI-Centre systematic review process

What is a systematic review?

A systematic review is a piece of research following standard methods and stages (see figure 1). A review seeks to bring together and ‘pool’ the findings of primary research to answer a particular review question, taking steps to reduce hidden bias and ‘error’ at all stages of the review. The review process is designed to ensure that the product is accountable, replicable, updateable and sustainable. The systematic review approach can be used to answer any kind of review question. Clarity is needed about the question, why it is being asked and by whom, and how it will be answered. The review is carried out by a review team/group. EPPI-Centre staff provide training, support and quality assurance to the review team.

Stages and procedures in a standard EPPI-Centre Review

- Formulate review question and develop protocol
- Define studies to be included with inclusion criteria
- Search for studies - a systematic search strategy including multiple sources is used
- Screen studies for inclusion
 - o Inclusion criteria should be specified in the review protocol
 - o All identified studies should be screened against the inclusion criteria
 - o The results of screening (number of studies excluded under each criterion) should be reported
- Describe studies (keywording and/or in-depth data extraction)
 - o Bibliographic and review management data on individual studies
 - o Descriptive information on each study
 - o The results or findings of each study
 - o Information necessary to assess the quality of the individual studies

At this stage the review question may be further focused and additional inclusion criteria applied to select studies for an ‘in-depth’ review.

- Assess study quality (and relevance)
 - o A judgement is made by the review team about the quality and relevance of studies included in the review
 - o The criteria used to make such judgements should be transparent and systematically applied
- Synthesise findings
 - o The results of individual studies are brought together to answer the review question(s)
 - o A variety of approaches can be used to synthesise the results. The approach used should be appropriate to the review question and studies in the review
 - o The review team interpret the findings and draw conclusions implications from them

Quality assurance (QA) can check the execution of the methods of the review, just as in primary research, such as:

- Internal QA: individual reviewer competence; moderation; double coding
- External QA: audit/editorial process; moderation; double coding
- Peer referee of: protocol; draft report; published report feedback
- Editorial function for report: by review specialist; peer review; non-peer review

The results of this systematic review are available in four formats:

SUMMARY

Explains the purpose of the review and the main messages from the research evidence

REPORT

Describes the background and the findings of the review(s) but without full technical details of the methods used

TECHNICAL REPORT

Includes the background, main findings, and full technical details of the review

DATABASES

Access to codings describing each research study included in the review

These can be downloaded or accessed at
<http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2416&language=en-US>

First produced in 2008 by:

Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square
London WC1H 0NR

Tel: +44 (0)20 7612 6367

<http://eppi.ioe.ac.uk/>

<http://www.ioe.ac.uk/ssru/>

The **Evidence for Policy and Practice Information and Co-ordinating Centre** (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

The views expressed in this work are those of the authors and do not necessarily reflect the views of the funder. All errors and omissions remain those of the authors.

This document is available in a range of accessible formats including large print. Please contact the Institute of Education for assistance:

telephone: +44 (0)20 7947 9556 email: info@ioe.ac.uk