

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

Review conducted by the Assessment and Learning Research Synthesis Group

Technical 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

EPPI-Centre report no. 1614 · October 2008

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.

© Copyright

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	4
1.1 Aims and rationale	4
1.2 Definitional and conceptual issues	5
1.3 Policy and practice background	6
1.4 Research background	7
1.5 Authors, funders and other users of the review	8
1.6 Review questions	8
2. Methods used in the review	9
2.1 User involvement	9
2.2 Identifying and describing studies	9
2.3 In-depth review	11
3. Identifying and describing studies: results	14
3.1 Studies included from searching and screening	14
3.2 Characteristics of the included studies (systematic map)	16
3.3 Identifying and describing studies: quality-assurance results	17
3.4 Summary of results of map	17
4. In-depth review: results	18
4.1 Further details of the studies included in the in-depth review	18
4.2 The impact of secondary school pupils of self and peer assessment	20
4.3 In-depth review: quality assurance results and possible weaknesses	27
4.4 Summary of results of synthesis	27
5. Implications	28
5.1 Strengths and limitations of this report	28
5.2 Implications for policy, practice and research	28
6. References	30
6.1 Studies included in map and synthesis	30
6.2 Other references used in the text of the Technical Report	32
6. Appendices	34
Appendix 1.1 Authorship of this review	34
Appendix 2.1 Inclusion and exclusion criteria	37
Appendix 2.2 Search strategy for electronic databases	38
Appendix 2.3 Journals handsearched	41
Appendix 2.4 EPPI-Centre keyword sheet v0.9.5	43
Appendix 3.1 Details of studies included in the systematic map	45
Appendix 4.1 Details of studies included in the in-depth review	47

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

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.

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
- 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
- The impact of staff development in self and peer assessment for the school workforce

CHAPTER ONE

Background

This chapter begins by outlining the aims of the review, and its rationale. This is the fifth review conducted by the Assessment and Learning Research Synthesis Group and it was commissioned by the Department for Education and Skills. In a policy context which is shaped by notions of ‘personalised learning’, the role of self and peer assessment is an important one.

1.1 Aims and rationale

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 how any impact depends on particular circumstances was sought so that, where trustworthy evidence was found, implications for policy and practice were 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). Modern theories of learning emphasise the active role of learners in constructing meaning from their experiences in order to develop understanding, and the active participation of learners in the processes of learning.

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. Without a serious focus on students’ ownership of their own learning processes, there is always the danger that the focus will be on curriculum delivery and teacher strategies which are less likely

to stimulate the sorts of intrinsic motivation for learning which is necessary for life in the twenty-first century. Indeed Flutter and Rudduck (2004) argued that in spite of decades of educational reform, pupils 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. Students learn these things through reflection and meta-cognitive development (Hattie et al. 1996, Black and Wiliam 1998a & b); they cannot be achieved by direct teaching. It requires a relationship between teachers and students in which students are helped to take responsibility for their learning and a view of learning that places learners at the centre of the process. It follows that the more learners know about, and participate in, decisions about the goals of their own learning, about where they have reached in relation to those goals and what further needs to be done to pursue them, the more they can direct their own learning efforts effectively. The potential role of self and peer assessment in achieving goals of education regarded as essential in preparing students to adapt to the accelerating changes in types of occupation and ways of living make it a key feature of educational practice.

1.2 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 *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. While all students have capacities for all of these outcomes, knowledge, skills and understanding are frequently the key focus of the subjects of the curriculum, their attainment being recognised by summative testing and assessment. On the other hand, the more difficult to assess values, attitudes and dispositions are often seen as the preserve of pastoral development and personal and social education.

These twin purposes of attainment and personal development come together in other well-recognised desirable student outcomes. For example, to become an effective learner there are particular values, attitudes and dispositions - such as curiosity, or meaning-making - which are elements of personal development necessary for success across the curriculum. Skills for enterprise, such as creativity and problem-solving and the values, attitudes and dispositions necessary for active citizenship, are another example of 'softer outcomes' that are integral to sustained achievement. 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.

Theoretically, it is possible for students to have low levels of attainment and high levels of personal development, or low levels of personal development and high levels of attainment. The most desirable combination of student outcomes would be high levels of both, where personal development and attainment are integrated.

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, as the term suggests, means that students make judgements about their own achievement and learning processes and, in the context of formative assessment, take part in decisions about the action they need to take to make 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. To take action they also need commitment to achieving the goals.

Peer assessment involves students in assessing each other's work, again through reflection on the goals and what it means to achieve them. Peer assessment 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. Peer assessment has particular value in formative assessment since students ask each other questions they may be inhibited from asking their teacher and explain things to each other using familiar language.

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. This requires that they both understand and are committed to the standard of work indicated in the criteria. The difference from formative assessment is that this does not lead to further learning, often because it takes place at the end of a piece or unit of work.

However, 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.

Both self and peer assessment in the context of formative assessment mean that learners must have an understanding of the goals of their work and of the criteria used in assessing their achievements. The process leads to the recognition by the learners of what further steps need to be taken to reach a particular goal, and to action on the part of the learners, possibly with the help of the teacher, to take these steps.

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 and peer 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 and peer assessment may have made little or no progress, suggesting limited outcomes relating to attainment.

1.3 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 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. And while in England there are still substantial requirements on schools to achieve higher standards, as measured by national assessment criteria, the concept of ‘personalised learning’ has emerged as a major focus for schools. The Department for Children, Schools and Families (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 policy shifts attention from the content of what is to be learned to the process of learning itself, and thus draws attention to the person who is the learner and to teaching strategies that promote learning. 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).

The third component of personalised learning is curriculum entitlement and choice. It is widely recognised that skills for life in the twenty-first century include the processing of knowledge and knowledge transformation and creation rather than just the repetition of set pieces of knowledge and the accumulation of information. Student entitlement to a relevant and meaningful curriculum that is tailored to their own needs is a goal that also requires student ownership of their own learning and the capacity to make choices.

Meaningful curricula are curricula that matter to the learner concerned. There are some high profile curricular initiatives, such as the RSA’s Opening Minds project, in which the starting point for the learning is a place, object or artefact which is of interest and which matters to the learner (Royal Society of Arts 2005). This and other similar initiatives tend to provide integration across subjects and include a focus on processes of learning such as creativity, critical thinking and problem solving. Central to these initiatives is self and peer assessment of both the specialised knowledge content of the curriculum and of the processes of learning.

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.

These issues often present problems in practice, especially in secondary schools where the realities of timetabling encourage twin tracks for personal development and attainment, and where the practices of assessment for learning are expected

to co-exist within an accountability system based on high-stakes testing, school inspection and performance management. Pollard and James (2004) suggest that one of the challenges to the concept of personalised learning is whether it is authentic - is the focus really on learners or is it on top-down teaching strategies designed to deliver pre-designated outcomes? Understanding how school systems and teachers can empower students to take responsibility for their own learning processes and pathways is thus a key concern.

Such attention to the learner as a person requires joined-up thinking by key agencies in education and beyond, which collaboratively attend to the well-being and growth of the learner as a person in a community. 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 - at least ideally - makes self and peer assessment a critical issue for both policy and practice, because it builds upon student self awareness, student ownership of their own learning process and student 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.

1.4 Research background

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

Inevitably, there is considerable interest in whether levels of performance are raised by self and peer assessment. McDonald and Boud (2003), in what they claimed was a unique study of the introduction of self assessment across a range of subjects on a large scale, 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 in a range of subjects. 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 overall mean scores of the two groups, in favour of those trained in self assessment, with some variation in impact across subjects.

The effect size was greatest for business studies and the humanities and least for science subjects. Although there was no comparison between trial and control groups in terms of how the self assessment

influenced their work, the trial students responded positively to the self assessment training and suggested that it helped them in preparing for the examination.

Black et al. (2003) discuss 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. In their quantitative findings, they report larger effect sizes than McDonald and Boud. However, in the Black et al. study differences between trial and control classes extended beyond the practice of self and peer assessment and included other components of formative assessment.

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).

Reporting non-cognitive impact depends, in several studies, on self-reported success or change. An exception is the work by Schunk (1996). In a study looking at both the effect of different goal orientation and of self assessment, with fourth grade students learning mathematics, outcome measures were a goal orientation inventory, a self-efficacy scale and a skill test of mathematics learned. Self assessment was an overwhelming factor accounting for any differences, beyond those noted from the manipulation of goal orientation. Only when self assessment was held constant was a difference associated with goal orientation evident. Self-assessment was associated with solving more mathematical problems and higher levels of self-efficacy.

There is relevant research dealing with practices in self-regulation and student participation in learning, which include but are not restricted to experience of self- and peer assessment.

Research into the dispositions, values and attitudes necessary for effective lifelong learning supports the active participation of learners in their own learning processes. 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, 2005).

These ideas are echoed in the research-validated Learner-Centered Psychological Principles of the American Psychological Association, which provides a knowledge base for understanding that learning and motivation are natural processes that occur when the conditions and context of learning are supportive of individual learner needs, capacities, experiences and interests. These include cognitive and metacognitive factors, motivational and affective factors, developmental and social factors, and individual difference factors (APA Task Force on Psychology in Education 1993; APA Work Group of the Board of Educational Affairs 1997).

Learner-centeredness is also related to the beliefs, characteristics, dispositions, and practices of teachers and it can be evaluated through students' perceptions of their teachers' learner-centred practices. 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; and
- treat learners as co-creators and partners in the teaching and learning process.

1.5 Authors, funders, and other users of the review

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 was 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, director of previous ALRSG, reviews 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 as noted in section 2.1.

1.6 Review questions

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?

CHAPTER TWO

Methods used in the Review

This chapter deals with the methods of the review and provides the information necessary for it to be replicated. It presents results of the stages of searching, screening using inclusion and exclusion criteria, and the application of the generic EPPI-Centre and review-specific keywords.

2.1 User involvement

2.1.1 Approach and rationale

The potential users of this review include all involved with education. However, the review was concerned with matters relating to the impact of self and peer assessment on student outcomes which is of particular interest to those making decisions about policy at national, local, school and classroom level. Thus, the main focus was to inform policy-makers concerned with assessment and practitioners and their professional bodies. The direct involvement of users in the conduct of the review was through their membership of the Review Group.

The Assessment and Learning Research Synthesis Group (ALRSG) at the time the review was completed, included the following users: 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 (AAlA), 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).

2.1.2 Methods used

Users were involved in the review in the following ways:

- As members of the review group, 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.
- One user member of the review group keyworded two studies.
- They responded to the emerging findings to provide verification of possible interpretations.

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

Inclusion criteria

The search for 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 type 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. Both naturally occurring and researcher-manipulated evaluation study types were considered to be relevant. Designs included comparison of the experience of comparable classes with different experiences of self or peer assessment or comparison of the same groups before and after the introduction of self or peer assessment. Surveys of students' and teachers' perceptions of the impact of student self or peer assessment and case studies reporting experiences and impacts of involving students in assessing their work were also considered relevant.

Exclusion criteria

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.
- D: Not research (excluded if not empirical study of particular procedures of assessment by teachers; also excluded if only procedure development was 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.

The full set of the inclusion and the exclusion criteria used to define this systematic review is given in Appendix 2.1 and section 3.1 gives the results of applying the inclusion and exclusion criteria.

2.2.2 Identification of studies: Search strategy

Studies were identified through a combination of (i) a two-stage strategy, used for databases and citations in already identified reports, where there was no immediate screening, and (ii) a one-stage strategy, where handsearching or online full text searching allowed immediate screening.

Studies were identified from the following sources:

- Bibliographic databases including AEI, BEI, ERIC, MLA, and PsycINFO
- Searches screening on full texts of journal publishers' web pages including both current (e.g. in IngentaConnect) and archived journals (e.g. in JSTOR)
- Handsearching of key journals in education (see Appendix 2.3)
- Citation searches of key authors and papers using Social Sciences Citation Index of ISI Web of Knowledge, and the British Library's Electronic Table of Contents (ZETOC)
- Reference lists of key authors and papers
- References on key specialist websites such as NFER, CRESST and SCORE
- Personal contacts and direct requests to key researchers in the area of self and peer assessment

Searches of these sources were limited so as to identify studies conducted in a specific time period of 1980-2005. The 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.

The terms used in the search of bibliographic databases is given in appendix 2.2. These were used as 'free text' in the search.

Different strategies were used, depending on the scope and structure of the databases and the number of hits on the first search attempts. If the first attempts using the four key terms 'self assessment', 'self evaluation', 'peer assessment' and 'peer evaluation' in combination with 'student' respectively produced only a small number of hits from a particular database or databases, then no further refining search strategies were applied. (Some database providers allow cross-database searching - for example, AEI and BEI can be searched together using the same search terms, as can MLA and PsycINFO.) However, if the first search attempts using the above four key terms and 'student' produced a large number of hits, we further refined our search strategies by adding more search terms listed in the second and/or third column of the table above. For example:

- In AEI and BEI, the following combinations of terms were used: 'student' and 'self assessment' (N=40), 'student' and 'self evaluation' (N=160), 'student' and 'peer assessment' (N=45), and 'student' and 'peer evaluation' (N=99). Using 'or' strategy to combine the four searches

above produced 240 hits. Therefore no further refinements were made.

- However, when the same combinations of terms were used in ERIC (1990-September 2004), they produced 788 hits for 'student' and 'self assessment', 2792 for 'student' and 'self evaluation', 103 for 'student' and 'peer assessment', and 582 for 'student' and 'peer evaluation'. Using 'or' strategy to combine the four searches above still produced a large number of hits (N=3,644). Therefore, we further refined our search strategy by adding the term 'school', resulting in 1,394 hits. Another level of refinement was added by using the terms in the third column of the table above: 'attitude' (N=51), 'motivation' (N=135), 'value' (N=72), 'responsibility' (N=115), 'disposition' (N=2), 'personal development' (N=6), 'achievement' (N=321), 'attainment' (N=19), 'learning outcomes' (N=17), 'study skills' (N=27), 'learning style' (N=10), 'self directed learning' (N=10), 'metacognition' (N=40), 'learning power' (Nil), or 'emotional literacy' (Nil). Using 'or' strategy, these produced a manageable number of hits (N=585).
- Subsequently, we also used the verbal phrases - e.g. 'self assess', 'self evaluate', 'self marking', 'self grading', 'self rating', 'self assessing', 'peer assess', and 'peer evaluate' to broaden the search scope in the AEI, BEI and ERIC. However, in MLA and PsycINFO, we used wildcards of these phrases such as 'self assess*', 'peer evaluat**', as this function is available in the two databases.
- In the British Library's ZETOC and the SSCI of ISI Web of Knowledge, the four key terms were searched as terms everywhere, because of the limited information pertaining to each bibliographic item in these two service providers.

The search logs were recorded. A database system was set up using the reference management programme EndNote for keeping track of and for coding studies found during the review. Titles, abstracts, and other relevant information were imported straight from the bibliographic databases, citation index search facilities and journal web pages if citation export systems were available, or were entered manually into our first EndNote database if not. Each paper in the first EndNote database was further labelled with the method of identification (e.g. database searching, handsearching). It was inevitable that there were duplicates because the searches were made in different databases and from different service providers. These were expunged.

2.2.3 Screening studies: applying inclusion and exclusion criteria

Screening of titles and abstracts entered into the database was carried out by the review team in order to check that they all met the inclusion

criteria. Studies were excluded if they met any of the exclusion criteria (see 2.2.1). Each excluded paper was labelled with the reason(s) for exclusion. Those papers judged as meeting the inclusion criteria were entered into a second database. Full reports were obtained, where possible, for these papers and the same inclusion and exclusion criteria were re-applied to the full text; those not meeting the inclusion criteria were excluded and labelled according to the reason(s).

2.2.4 Characterising included studies

The included studies were keyworded using EPPI-Centre Core Keywording Strategy for Classifying Education Research (EPPI-Centre 2003). Additional keywords specific to the context of this systematic review, with guidelines for application, were added to those of the EPPI-Centre. Both the generic and specific keywords (see Appendices 2.4 and 2.5) were applied to the included studies. The included studies were allocated to the review team member in such a way that each paper was keyworded independently by two people. Keywording of all included studies for which it was possible to obtain full texts before the end of November 2005 was carried out, using both the generic and specific keywords.

Keywording resulted in the exclusion of some studies that were found not to meet the inclusion criteria on close reading of the text. The agreed keywords for the remaining studies were used to produce the systematic map of the included studies. All the keyworded studies have been added to the larger EPPI-Centre database, REEL, for others to access via the website.

2.2.5 Identifying and describing studies: quality assurance process

Records were made of all searches: electronic database searches were documented, and dates of journals handsearched were recorded. Application of the inclusion and exclusion criteria and the keywording was conducted by pairs of review group members working independently and then comparing their decisions and coming to consensus through discussions. Members of the EPPI-Centre also 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).

2.3 In-depth review

2.3.1 Moving from broad characterisation (mapping) to in-depth review

The included studies were described ('mapped') in terms of the keywords and various tables presented to a meeting of the Review Group. It was decided that all 51 studies remaining after keywording

appeared to be equally relevant to the review question and should be included in the in-depth data extraction. However, due to the workload of data extracting all the included studies, several options for using a proportion of the 51 studies were proposed to the DfES. It was agreed that the priority of data extraction for this systematic review should be on the studies that were conducted in secondary education contexts where the issues are arguably more complex than in primary schools and where students would be more articulate about self and peer assessment. Therefore, a sub-group of 26 studies conducted in contexts of secondary education were included for in-depth review. The remaining 25 studies of primary education are listed in the references in section 6.

2.3.2 Detailed description of studies in the in-depth review

The 26 keyworded studies of secondary education were entered into the EPPI-Centre's detailed data extraction software, EPPI-Reviewer, and data extracted using EPPI-Centre generic (EPPI-Centre 2003) and review-specific questions relating to the weight of evidence of each study in the context of this review.

2.3.3 Assessing quality of studies and weight of evidence for the review question

In order to ensure that conclusions were based on the most sound and relevant evidence, judgements were made using the EPPI-Centre 'weight of evidence' criteria. This involved judgements about 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. These answers were given on the basis of the information in the study report, which may or may not have given an account of all aspects of the study required for judging the soundness of the research. The judgement of methodological soundness was thus dependent on what was reported in the study by the authors. The lack of information about a certain feature did not necessarily mean that this feature was not attended to in practice by the study. Studies were rated as high, medium or low in relation to methodological soundness, according to what was reported. This judgement was not review-specific.

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

The second 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. In theory, some study types or designs might be better matched than others to the focus of the review. 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. Studies were rated high, medium and low in relation to this aspect.

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

As in B, this judgement concerned 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 the topic focus (including conceptual focus, context, sample and measures) of the study: that is, how well the focus of the data collected helped to answer the review questions. Again, the judgements were review-specific and made in terms of high, medium or low relevance.

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. The rationale for this was that a study judged to be giving evidence of only medium weight on account of relevance of focus, context, sample and measures could not provide high weight of evidence overall for the review.

2.3.4 Synthesis of evidence

The data were synthesised to bring together the studies which answered the review questions and which met the quality criteria relating to appropriateness and methodology. The structure for the synthesis of evidence from the in-depth review was taken from the review question, 'What is the evidence of the impact on students in secondary schools of self and peer assessment?' and its sub-questions: (a) 'How does any impact vary with the characteristics of the students and the approaches used in self and/or peer assessment?' (b) 'What conditions affect the impact of self and peer assessment?' and (c) 'What are the implications for assessment policy and practice of these findings?'

The weight of evidence assessments and the responses to the generic and specific review questions in the data extraction were then used as a basis for producing a narrative synthesis to address the review questions. A summary of the characteristics of the included studies is presented in Appendix 4.1 and further details of the included studies can be found in the uploaded data extraction, available on the EPPI-Centre database REEL, which can be accessed via the EPPI-Centre website.

2.3.5 In-depth review: quality assurance process

All in-depth data extraction was also carried out independently by at least two people, as in keywording, using the generic data extraction and quality assessing guidelines on EPPI-Reviewer and review-specific questions. The review team members and one EPPI-Centre member of staff were allocated studies for data extraction and quality assessing in such a way that the keywording and data extraction for each study involved different people. 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 the EPPI-Centre staff for quality-assurance purposes and again any differences were resolved by discussion.

CHAPTER THREE

Identifying and describing studies: results

This chapter presents results of the stages of searching, screening using inclusion and exclusion criteria, and the application of the generic EPPI-Centre and review-specific keywords. The numbers of studies at the various stages of filtering studies are given in a flow diagram. The characterisation of the selected studies in terms of the keywords is described and the results are given of the quality-assurance procedures for these parts of the process.

3.1 Studies included from searching and screening

The sources of the initial papers found and, for comparison, the sources of the studies that were included in keywording are given in Appendix 3.1.

The number of papers and studies at different points in the searching and screening processes are summarised in Figure 3.1 below. It can be seen that the total number of papers screened was 214.

The criteria for excluding papers and the number excluded at all stages are given in Table 3.1. There were 158 papers excluded, some being excluded for more than one reason. Two papers were identical, but the author used two different names, perhaps because of marriage.

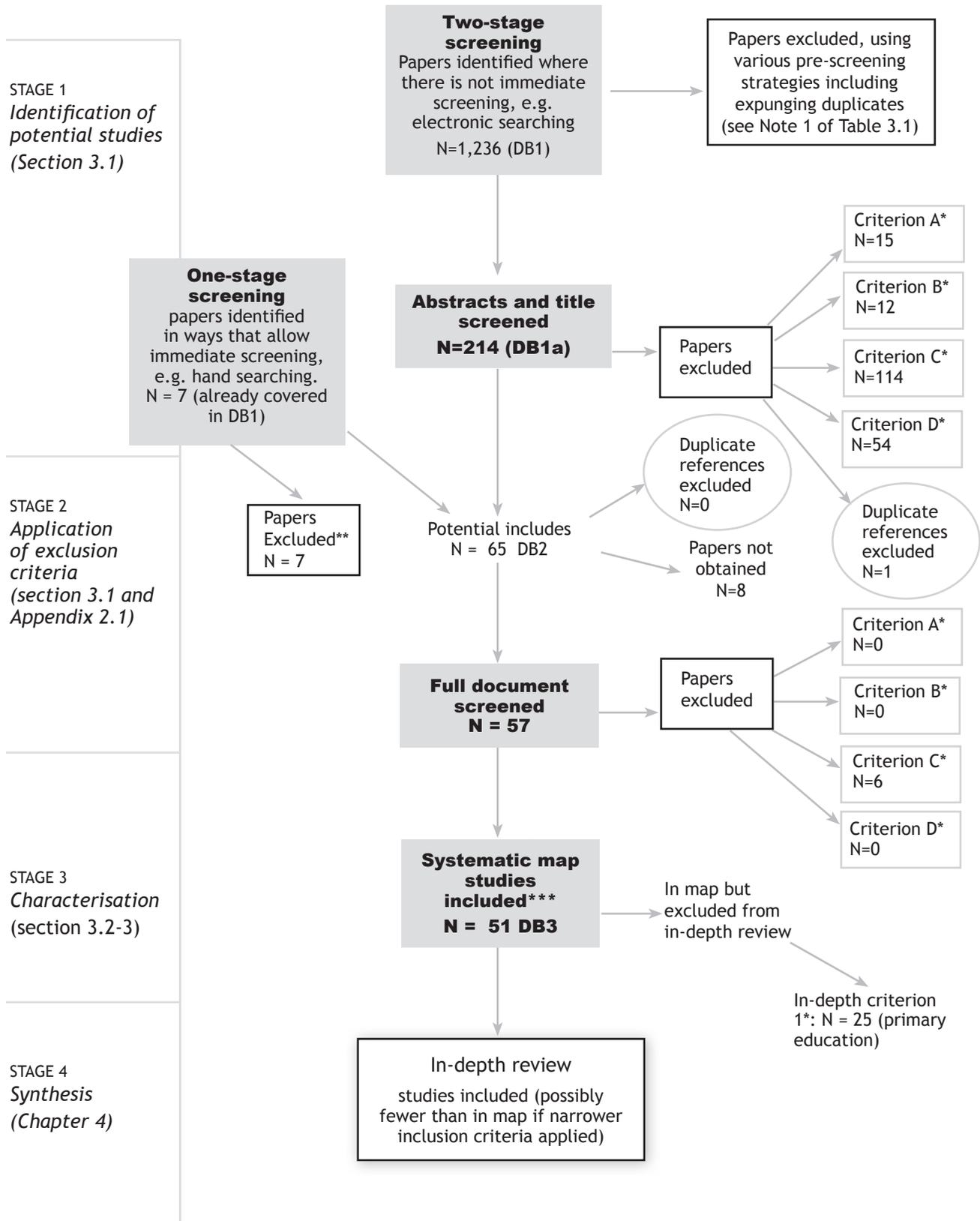
One was then excluded from the data extraction as a separate item so that only one of these appeared in the list of studies used in data extraction.

In the screening process, all papers were labelled either IN or OUT with the reason(s) for exclusion. In addition, some papers, considered to be of particular relevance but excluded for one of these reasons, were labelled as useful for background discussion. Of the 65 papers labelled IN, the full texts of 8 could not be obtained in the timescale, leaving 57 for the keywording stage. At this stage, 6 further studies were excluded according to exclusion criteria C (i.e. not reporting impact). Thus 51 studies remained in the systematic map and 26 of these were identified for in-depth review (see 2.3.1).

Table 3.1 Exclusion criteria and numbers excluded at all stages (not mutually exclusive)

Exclusion criteria	Number of studies excluded
A: Not self or peer assessment (excluded if students had no part in collecting and interpreting information about their performance)	15
B: Not related to education in school (excluded if studies were related to college students; higher education; nursing education, other vocational)	12
C: Not reporting impact on students of the process of self or peer assessment rather than just the outcome of the assessment	114
D: Not research (excluded if not empirical study of particular procedures of assessment by teachers; also excluded if only procedure development was reported or description without report of use; excluded if handbooks, textbooks and reviews)	54

Figure 3.1 Filtering of papers from searching to map to synthesis



Key for Figure 3.1

*Criteria for exclusion are not mutually exclusive, but were applied sequentially so that the data were mutually exclusive.

3.2 Characteristics of the included studies (Systematic Map)

This section describes the characteristics of the 51 studies that were keyworded and included in the systematic map.

3.2.1 Characterisation in terms of the EPPI-Centre keywords

Appendix 3.1 shows the countries of the studies keyworded. The studies were mainly from the US, with English-speaking countries totalling 45 of the 51 studies keyworded. Given that the report being written in the English language was specified as an inclusion criterion this is unsurprising. However, the dominance of studies that took place in North America raises important issues regarding the potential transfer of findings, since the educational context in general, and assessment context in particular, is significantly different.

Table 3.2 below shows that the studies keyworded were nearly all 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.

Table 3.2 Which type(s) of study does this report describe? (not mutually exclusive, N=51)

Attribute	Number
Description	3
Exploration of relationships	20
Evaluation: naturally occurring	3
Evaluation: researcher-manipulated	47

3.2.2 Characterisation in terms of the review-specific keywords

Table 3.3 below shows that more than half of the studies keyworded focused on formative assessment (though some studies were not categorised as either formative or summative, with details given). Furthermore, as anticipated given the focus of the review, all were characterised by one or more forms of self and/or peer assessment.

Table 3.3 What types of assessment does the report describe? (not mutually exclusive, N=51)

Types of assessment	Number
Summative	10
Formative	29
Portfolio	8
Personal report	2
Journal	2
Other	12
Self or peer marking/grading	18
Self or peer written comments	11
Self or peer oral comments	13

Table 3.4 shows the outcomes of assessment reported upon in the studies keyworded. 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. Given the emphasis on improving learning-to-learn skills in the literature on self and peer assessment, this focus on engagement as an outcome might be expected. 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 those that have received much greater attention in recent years - 'well-being' as part of the *Every Child Matters* agenda and pupils' understanding of progression, which has emerged more recently as an issue in relation to progress in the national curriculum. This might reflect the large proportion of studies in the systematic map that focused on North America and were published before 2000. In addition, 'well-being' is an outcome that presents considerable challenges in terms of measurement.

Table 3.4 What outcomes of assessment are described in the study? (not mutually exclusive, N=51)

Outcomes of assessment	Number
Achievement	37
Enjoyment	7
Engagement in learning	22
Social engagement	6
Confidence to participate in learning community	9
Well-being	1
Pupils' understanding of progression within the subject	5
Other	8

The subject foci of the studies keyworded are given in Appendix 3.1 Evidence of subject-specific differences was of interest in the review. Given the focus on learning skills, 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 through consultation adopt the goals and criteria identified by the teacher
- **High:** where students determine the goals and are committed to engaging in learning to achieve them

Table 3.5 shows the levels of ownership for the studies keyworded. Despite the review focusing on self and peer assessment which might be assumed to be associated with high levels of pupil ownership of the process of assessment, the studies keyworded were evenly distributed across all three levels of ownership. It has often been suggested in the literature (e.g. Flutter and Rudduck 2004), that student commitment to engaging in learning increases with higher levels of ownership.

Table 3.5 Level of ownership by pupils of the assessment process

Level of ownership	Low	Medium	High	Total
Numbers of studies	19	14	18	51

In the 51 studies keyworded, higher levels of student ownership were found where both self and peer assessment approaches were used as shown in Table 3.6. 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.

3.3 Identifying and describing studies: quality assurance results

The review-specific keywords used to identify and describe these studies were produced as a result of one review team meeting involving the user members followed by an email consultation. All of the 51 studies were keyworded by two people independently, moderated and reconciled if there was a difference, by face-to-face or telephone contact. The agreed set of keywords were entered into EPPI-Reviewer (EPPI-Centre 2003) with the

generic key words, and the results for each study were entered for both sets of key words. The reports were then generated through the EPPI-Reviewer's reporting facility. The final map was approved by both the review team and the review group.

Table 3.6 The relationship between type of assessment and level of ownership (not mutually exclusive, N=51)

Type of assessment / ownership	Low	Medium	High
Summative	4	4	2
Formative	7	9	13
Portfolio	1	2	5
Personal report	0	2	0
Journal	1	1	0
Other (please specify)	5	5	2
Self or peer marking/grading	3	5	10
Self or peer written comments	2	3	6
Self or peer oral comments	4	3	6

3.4 Summary of results of 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 pupils 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

In depth review: results

The in-depth review question was:

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

Further questions to be addressed by the in-depth review 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?

4.1 Further details of the studies included in the in-depth review

The outcomes of the searching and keywording reported in section 3 confirm that there was no shortage of studies in this area. The in-depth review focused only on the 26 studies undertaken with secondary aged pupils. The characteristics of the pupils, such as ethnicity, gender, prior attainment and presence of special educational needs, are reported below specifically, as they relate to one of the research questions. Similarly, the curricular areas, type of assessment involved and types of outcomes reported are shown in Table 4.4 below as they relate to the key research questions.

More than half (16) of the 26 studies in the in-depth review were conducted in the US, 3 in England (one of which was also conducted in Australia), 2 in Canada, 2 in Latvia and one each in Barbados, South Africa and Finland. The implications of this wide spread of contexts is commented upon in section 5.

Nearly all the studies included in the in-depth review were researcher-manipulated evaluations.

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

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. 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.

Table 4.2 Systematic review of the evidence of the impact on students in secondary schools of self and peer assessment. Weight of Evidence

Item	A: Taking account of all quality assessment issues, trustworthiness of the study findings in answering the study questions.	B: Appropriateness of research design and analysis for addressing the questions, of this specific systematic review.	C: Relevance of particular focus of the study for addressing the questions of this specific systematic review.	D: Taking into account A, B and C, overall weight of evidence this study provides for addressing this systematic review question.
Bickmore (1981)	Medium trustworthiness	Medium	Medium	Medium
Brookhart (2001)	Medium trustworthiness	Medium	High	Medium
Bruce (2001)	High trustworthiness	High	High	High
Carter (1997)	Low trustworthiness	Low	Low	Low
Crouch et al. (1997)	Medium trustworthiness	Medium	Medium	Medium
Deakin Crick et al. (unpublished)	Medium trustworthiness	High	Medium	Medium
Ezell et al. (1999)	Medium trustworthiness	High	Medium	Medium
Goodrich (1997)	High trustworthiness	High	High	High
Gregait et al. (1997)	Low trustworthiness	Medium	Low	Low
Hewitt (2002)	Medium trustworthiness	Medium	Medium	Medium
Katstra et al. (1987)	Medium trustworthiness	High	High	High
Kitsantas et al. (2004)	High trustworthiness	High	Medium	Medium
Klein (1998)	Medium trustworthiness	Medium	Low	Low
Klenowski (1995)	Medium trustworthiness	Medium	High	Medium
Knubb-Manninen (1994)	High trustworthiness	Medium	Low	Low
Marshall (1993)	High trustworthiness	High	High	High
Masqud and Pillai (1991)	High trustworthiness	High	High	High
McDonald and Boud (2003)	Medium trustworthiness	High	High	High
Olina and Sullivan (2002)	High trustworthiness	High	High	High
Olina and Sullivan (2004)	High trustworthiness	High	High	High
Powell and Makin (1994)	Low trustworthiness	Medium	Low	Low
Rief (1990)	Low trustworthiness	Low	Medium	Low
Ross et al. (1998)	Medium trustworthiness	Medium	High	Medium
Ross et al. (2002)	High trustworthiness	Medium	Medium	Medium
Uselman (1996)	Low trustworthiness	Low	Medium	Low
Young et al. (1997)	Medium trustworthiness	Medium	Medium	Medium

Table 4.2 shows that the overall weight of evidence was high for 8 studies, medium for 11 studies and low for 7 studies. The studies that were rated low overall weight of evidence are included in the synthesis as they have specific contributions to make in addressing the research questions, but they are given less weight in the conclusions drawn. Carter (1997) makes a contribution in looking specifically at pupils designated as gifted. Gregait et al. (1997) utilised journals to promote self assessment and was one of only two studies (the other was Crouch et al. 1997) that focused on physical education, thus providing insights relating to practical skills. Klein (1998) is a meticulously designed study that adopts a limited view of self monitoring which reduces its capacity to address the review question but it does involve a high level of ownership by students in the assessment process. Similarly, Knubb-Manninen (1994) is both well designed and executed and provides a notable example of how students do not self evaluate, thereby reporting little about its impact. Both Powell and Makin (1994) and Rief (1990) are small scale studies focusing on interventions in one class within a school. Powell and Makin is one of only two studies included that focuses on students with special educational needs and Rief provides useful insights into the process of increasing pupil engagement. Uselman (1996) specifically addresses ownership of the assessment process and provides a strong focus on peer evaluation.

4.2 The impact of secondary school pupils of self and peer assessment

The findings address each review question in turn.

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

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). In order to address the review questions, the findings are related to the specific characteristics of pupils, evidence of subject specific differences, any variation in impact relating to the different approaches used in self and/or peer assessment and the conditions which affect the impact of self and peer assessment.

4.2.1 Increased pupil attainment

Fifteen of the twenty-six studies reported on performance outcomes. Nine reported some increase in pupil attainment, though in one of these studies (Knubb-Manninen 1994, low WoE) this impact was weak. Knubb-Manninen suggests that the lack of strong evidence arose from pupils being insufficiently aware of their own learning processes and habits to realise that they could uncover

possible differences by self-evaluation. Four of the nine studies (Knubb-Manninen 1994, Powell and Makin 1994, Rief 1990, Uselman 1996) that reported an increase in attainment 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.

Examples of improved attainment included Bickmore (1981, medium WoE) who noted that students in the self evaluation group improved language scores significantly more than those in the control group, a finding supported by teacher perceptions. Masqud and Pillai (1991, high WoE) similarly reported from a randomly controlled trial that self evaluation improved subsequent test performance in science. Crouch et al (1997, medium WoE) found that peer assessment in volley ball led to improved ball skills and Goodrich (1997, high WoE) and McDonald and Boud (2003, high WoE) both reported that students who were taught to assess themselves showed higher pre- to post-test gains on content learning than the students in their control groups.

Five studies (four rated high, one medium WoE) reported no significant increase in attainment or performance. 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 (Ross et al. 1998, medium WoE; Olina and Sullivan 2004, high WoE) found no differences in attainment following self evaluation, Ross et al. on mathematical probability tests and Olina and Sullivan on an independent rater’s evaluation of the students’ written project reports. Both Olina and Sullivan (2002, high WoE) using measures of independent project evaluation and Hewitt (2002, medium WoE) drawing on ‘expert’ assessments of musical performance noted that evaluations improved performance whether done by the teacher or the student. Olina and Sullivan noted highest grades on work at post-test for students evaluated by the teacher only.

4.2.2 Improvements in self esteem

In keeping with previous research (e.g. Marsh et al. 2005), self esteem was accepted in a number of 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 a student completed questionnaire (Bickmore 1981, medium WoE; Klenowski 1995, medium WoE; Marshall 1993, high WoE; Olina and Sullivan 2002, 2004, both high WoE) while three (Carter 1997, low WoE; Deakin Crick et al. unpublished medium WoE; Kitsantas et al. 2004, medium WoE;) used self reporting by students through interviews and in Ezell et al. (1999, medium WoE) teachers and parents were also interviewed for their perceptions of the students’ self esteem. Seven of these studies (one low, five medium, one high WoE) reported positive

outcomes. In the eighth study (Marshall 1993), only a small minority of students were reported to have improved on self esteem and in the ninth study (Olina and Sullivan 2002) the control group achieved higher scores than the intervention group.

4.2.3 Improvements in learning to learn

Twenty of the twenty-six studies reported on outcomes relating to aspects of 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 rather than observation or 'test' measures, though Deakin Crick et al.'s (unpublished, medium WoE) ELLI begins to engage with this in a more 'objective' way.

Seventeen (five high, eight medium, four low WoE) of the twenty studies reported positive outcomes on goal setting, clarification of objectives, increased responsibility for learning and/or increased confidence. 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. Transfer of learning was also noted as a positive outcome by Olina and Sullivan (2002, 2004, both high WoE). Bruce (2001, high WoE) concluded that student engagement increased and students took more responsibility for their own learning including sharing the responsibility with the teacher for identifying the criteria for evaluation. Student engagement and motivation was also noted to increase in the studies by Ezell et al. (1999, medium WoE), Rief (1990, low WoE) and Young et al. (1997, medium WoE).

Four studies (Bickmore 1981, medium WoE; Gregait et al. 1997, low WoE; Marshall 1993, high WoE; Ross et al. 1998, medium WoE) reported improved study skills in particular, relating to setting goals and clarifying expectations. 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 (Goodrich 1997, Hewitt 2002, Klein 1998) concluded that learning to learn and/or metacognition did not increase following self assessment. Goodrich (1997, high WoE) found a link between self assessment and metacognition for girls but no clear link between metacognition and learning. This study noted that students who assessed their own work were remarkably willing to revise it and concluded that although the

cognitive mechanisms underlying self assessment are still in question, self assessment can increase self-regulation in some students and learning in many. Hewitt (2002, medium WoE) reported that self guided evaluation forms did not increase students' ability to self assess accurately in musical performance and concluded that self assessment skills specifically have to be taught. Klein (1998, low WoE) noted that self monitoring did not support metacognitive activity or affect the ability to transfer learning though as noted elsewhere, this study adopts a more limited view of self assessment than that defined in this review.

Many studies reported that dialogue about learning increased. Ezell et al. (1999, medium WoE), evaluating the use of portfolios with pupils who had learning disabilities, noted improved teacher-pupil, teacher-parent and pupil-parent dialogue. In another study (Ross et al. 2003, medium WoE), students' judgements about their performance were informed by parental interpretations, but these interpretations were based on data alone without seeing the work itself, which led to overly narrow views. The researchers concluded that parents were unable to interpret alternative assessments without greater evidence and support from teachers than was provided.

4.2.4 Characteristics of the pupils

Eleven studies reported on the ethnicity of the students involved. Of these five (Bruce 2001, Goodrich 1997, Gregait et al. 1997, Katstra et al. 1987, Uselman 1996) focused on student populations who were mainly or all white. One study (Marshall 1993) simply described the population as 'mixed'. The other five studies reviewed had samples in which a high proportion of students were from other ethnic groups, two undertaken in Latvia (Olina and Sullivan, 2002, 2004) with exclusively Latvian students, and one (Klein 1998) in the US with 98 percent students from Asian-American, African-American or Latino students. In the other two studies (Kitsantas et al. 2004, Ross et al. 2002), just over half of the students were white with the others from a variety of ethnic groups.

In 25 of the studies the population was mixed boys and girls. One (Marshall 1993) was undertaken in an all girls' school. Two studies (Ezell et al. 1999, Powell and Makin 1994) focused on pupils identified as having special educational needs, one study sample (Carter 1997) was of pupils designated gifted and talented, and one (Young et al. 1997) described the sample as 'disenchanted'.

No significant differences were found on the impact of self and peer assessment for different groups of students according to gender, ethnicity or prior attainment. One study (Goodrich 1997, high WoE) demonstrated that criterion-referenced assessment had a positive effect on girls and negative effect on boys. Another study (Katstra et al. 1987, high WoE) showed that boys' writing performance was better when they were taught by a male teacher.

The outcomes for pupils with special educational needs focused on increased self esteem. One study (Ross et al. 2003) concluded that older students were more effective self evaluators and that an implication of this was that training in self and peer assessment should start at an earlier age. Twenty-six of the studies that were keyworded focused on primary/elementary schools but were excluded from in-depth review when it was decided to focus this review on secondary pupils.

The studies provided only limited comparative outcome data of different ‘ability’ (usually defined as such by prior attainment), or ethnic groups, though studies that looked at these characteristics, for example, Knubb-Manninen (1994), noted no differences that could be attributed to ability or second language learners. Individual studies suggest self and peer assessment can be effective with pupils identified as gifted and talented and those having SEN and with different ethnic groups.

4.2.5 Evidence of subject specific differences

Six studies (Bickmore 1981, Brookhart 2001, Katstra et al. 1987, Marshall 1993, Rief 1990, Young et al. 1997) focused on literacy or English, four (Brookhart 2001, Goodrich 1997, Klenowski 1995, Masqud and Pillai 1991) on science, three (Carter 1997, Powell and Makin 1994, Ross et al. 1998) on mathematics, two (Crouch et al. 1997, Gregait et al. 1997) on physical education and one each on modern foreign languages (Knubb-Manninen 1994), ICT/media (Kitsantas et al. 2004) and music (Hewitt 2002). Five had no particular curricular focus or didn’t specify a subject and four were cross-curricular.

No clear findings emerged on differences between subjects, since only one of the 26 studies compared self and peer assessment across different subjects. McDonald and Boud (2003) 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 overall mean scores of the two groups, in favour of those trained in self assessment, with some variation in impact across subjects. The effect size was greatest for business studies and the humanities and least for science subjects.

Comparison across studies was limited by the intervening variables of characteristics of students, settings and type of assessment involved. But the two studies focusing on physical education (Crouch et al. 1997, medium WoE; Gregait et al. 1997, low WoE) suggest that practical skills and thereby subjects may be more receptive to immediate changes than more ‘academic’ ones. These subjects may perhaps be less receptive to establishing the

deep changes in learning to learn that might transfer across subjects as has been observed for example, in assessment for learning research (Black et al, 2003). Conversely, one study (Ross et al. 1998, medium WoE) concluded that mathematics was more difficult to self assess as the terms and concepts are more challenging to understand.

Black et al. (2003) discuss how differences among subject disciplines may affect how teaching and learning take place and may account for some differences in impact of attempts to foster self-regulation through self and peer assessment. In their quantitative findings they report significant effect sizes. However, the differences between trial and control classes in their study extended beyond the practice of self and peer assessment and included other components of formative assessment.

4.2.6 How does the impact vary with the approaches used in self and peer assessment and what are the differences between self and peer assessment?

More than half the studies focused on formative assessment as noted in Table 4.3 below, though some that were not stated to be either about formative or summative assessment addressed both, for example by focusing on goal setting or teacher grading but with a formative component such as dialogue between teacher and student. While self and peer assessment are more usually associated with formative assessment, some interventions involved student summative assessment, for example by inviting students to grade their own or others’ projects.

Table 4.3 What types of assessment were involved? (not mutually exclusive, N=26)

Attribute	Number
Summative	4
Formative	14
Portfolio	5
Journal	1
Other	5
Self or peer marking/grading	11
Self or peer written comments	6
Self or peer oral comments	6

Fifteen studies focused on peer assessment instead of, or as well as self assessment. Three studies (Bickmore 1981, medium WoE; Crouch et al. 1997, medium WoE; Powell and Makin 1994, low WoE) reported positive impact on attainment of self and peer assessment. Three studies (Hewitt 2002,

medium WoE; Katstra et al. 1987, high WoE; Ross et al. 1998, medium WoE) reported no significant increase in attainment from the use of self and peer assessment and two studies (Goodrich 1997, high WoE; Kitsantas et al. 2004, medium WoE) reported an increase from self assessment only. Five studies noted increases in self esteem and learning to learn and a further four reported improvements only in learning to learn. One study noted improvements in learning to learn from self assessment only and another from peer assessment only.

Hence, no clear differences emerged between the effects of self and peer assessment. Self assessment is sometimes assumed to be easier for students than peer assessment, as no dialogue necessarily is involved, though at some stage the student is required to discuss their evaluation with the teacher. However, some of the studies reviewed (e.g. Ross et al. 1998, medium WoE) suggested that self assessment is more demanding as the student needs to understand what and how to assess their work. In peer assessment, the quality of peer dialogue, in particular reasoning skills is crucial but students can learn the skills of what and how to assess from one another. Klenowski (1995, medium WoE) concluded that peer evaluation enhanced interaction and self evaluation, arguing that this enables students to identify teachers' tacit knowledge through comparing the assessment of their work by peers against given criteria to that undertaken by the teacher. If a teacher's evaluation differs from that reached through peer assessment, then students learn about the teachers' standards for the work and might enter a dialogue with the teacher about these.

4.2.7 What conditions affect the impact of self and peer assessment?

(a) Involvement of students in specifying criteria for evaluation

Around a third of the studies (e.g. Bruce 2001, high WoE; Gregait et al. 1997, low WoE; Klenowski 1995, medium WoE) suggest that the involvement of the students in 'co-designing' the criteria for evaluation is important in achieving positive outcomes. Knubb-Manninen (1994, low WoE) reported that students need to be aware of the targets they are trying to achieve in order for self-evaluation to be effective:

Evaluation of one's learning is a central metacognitive skill. If we want to facilitate students' ability of self-evaluation in order to make them capable of more autonomous learning, more attention must in future be given to the problem identification process. (p157-8)

Two studies (Kitsantas et al. 2004, Young et al. 1997, both medium WoE) suggest that for self evaluation to be effective, students need these targets to focus on outcome rather than process goals. This contradicts previous research by Schunk (1996) suggesting that self evaluation improved student achievements regardless of goal orientation.

Klenowski (1995, medium WoE) reported that positive outcomes for both self esteem and learning to learn occurred only when the focus was on the criteria for evaluation, not on the student:

The self-esteem of students was preserved when the focus was on the criteria for evaluation and not on the student. The provision of criteria assisted students to distance themselves from their peers, themselves, and to maintain an objective focus. (p155-156)

(b) Level of ownership by students

In the larger set of studies keyworded (51 in all), higher levels of student ownership were found where both self and peer assessment approaches were used. However, from the 26 studies data extracted, the relationship between positive outcomes and higher levels of ownership remains unclear (see table 4.4 above). Half the studies described high levels of ownership in the process and this emerged as related to aspects of pedagogy, rather than to specific subject areas. For example, Bruce (2001, high WoE) noted impact on the expectations of students when they were involved in clarifying and co-designing assessment criteria, but no specific (immediate) effect on subject knowledge.

(c) Culture

Social climate is noted in some studies reporting positive outcomes to be important and 'active listening' is one element of this. Pupils begin to behave in the same way in discussions with other pupils as that modelled by the teacher (Powell and Makin 1994, low WoE). While there is a vast literature on defining 'culture' which cannot be summarised here, a number of conditions emerged under which self and peer assessment seem to be more effective, which can best be described generically as relating to culture. These include the need for teacher commitment to learner control, developing a language for dialogue about learning (e.g. Deakin Crick et al. unpublished) and moving from a dependent to an interdependent relationship between teacher and students which enables teachers to respond pedagogically to student feedback (Klenowski 1995).

4.2.8 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 in Klenowski's (1995, medium WoE) study in which the importance of teacher and student defining the criteria together is emphasised. Other factors that emerge from these and other studies include students developing a better grasp of their own strengths and weaknesses, students becoming more accountable for their own learning, self and peer assessment providing a broader evaluation

Table 4.4 Weight of evidence, level of ownership, assessment approach, subject focus and outcomes

Study	D: overall weight of evidence	Level of ownership	Approach to assessment	Subject	Outcomes relating to attainment	Outcomes relating to self esteem	Outcomes relating to learning to learn
Bickmore (1981)	Medium	High	Self or peer marking	literacy second language	positive effect on literacy > control	Self concept improved > control	teachers rated increase in student capability
Brookhart (2001)	Medium	Low	Formative and summative self assessment	English and science			students' self assessment in approach to study, transfer of learning
Bruce (2001)	High	Medium	Formative self assessment	none			student engagement & responsibility for own learning increased
Carter (1997)	Low	High	Formative summative, self or peer marking	maths		increased motivation from personal feedback	
Crouch et al. (1997)	Medium	Medium	Self or peer marking	PE	improved ball skills		
Deakin Crick et al. (unpublished)	Medium	High	Formative, self or peer marking	none		resilience and strategic awareness scores increased	increased learning power
Ezell et al. (1999)	Medium	Low	Portfolio	cross-curricular	communication skills improved	increased self esteem in pupils with SEN	increased motivation
Goodrich (1997)	High	Medium	Formative, self or peer marking	science	self assessment increased subject learning		no increase in metacognition
Gregait et al. (1997)	Low	Low	Journal, self or peer written comments	PE			effective use of journals, self goal-setting led to greater focus
Hewitt (2002)	Medium	Low	Self or peer marking	music	feedback affected performance whether done by teacher or self		ability to self assess in music did not improve
Katstra et al. (1987)	High	Medium	Formative, self or peer oral comments	literacy	writing skills no better following self evaluation		peer evaluation improved writing attitudes for pupils of 2/3 teachers

Study	D: overall weight of evidence	Level of ownership	Approach to assessment	Subject	Outcomes relating to attainment	Outcomes relating to self esteem	Outcomes relating to learning to learn
Kitsantas et al. (2004)	Medium	Medium	Formative, self or peer marking	animation media	self-evaluation had positive effect on students who received outcome goals but not on those instructed to adopt a process goal	enhanced self efficacy and motivation to pursue tasks	
Klein (1998)	Low	High	Formative, self or peer written comments	none			
Klenowski (1995)	Medium	High	Formative, portfolio self and peer	science		increased self-awareness, confidence, understanding areas for improvement	
Knubb-Manninen (1994)	Low	High	Formative	modern languages	weak evidence that self evaluation is associated with successful learning		
Marshall (1993)	High	High	Formative	English		minority reported increase in self esteem	students reported increased study skills
Masqud and Pillai (1991)	High	Low	Summative	science	self evaluation improved subsequent test performance		
McDonald and Boud (2003)	High	High	Summative and self assessment	other curricular	higher academic outcomes > control from training in self assess; effect size greatest for business studies and humanities, least for science subjects		
Olina and Sullivan (2002)	High	Medium	Formative, summative, self assessment	cross-curricular	any feedback improved marks on work but teacher evaluation did so most	control group had most positive post-test attitudes	teacher plus self evaluation led to greater confidence and transfer

Study	D: overall weight of evidence	Level of ownership	Approach to assessment	Subject	Outcomes relating to attainment	Outcomes relating to self esteem	Outcomes relating to learning to learn
Olina and Sullivan (2004)	High	Low	Formative, summative, self assessment	cross-curricular	no differences between self evaluation and control groups	most positive attitudes in students in self evaluation groups	self evaluation led to greater confidence and transfer
Powell and Makin (1994)	Low	Low	Self or peer oral comments	maths	improvements in maths performance but authors cautious about attributing to self reflection		students responded to peer comments in their explanations. more supportive to each other over time
Rief (1990)	Low	High	Portfolio	literacy			engagement in learning improved through using portfolios
Ross et al. (1998)	Medium	High	Formative, self or peer marking	maths	self evaluation no impact on student achievement		self evaluation training clarifies expectations in maths
Ross et al. (2003)	Medium	High	Formative, self or peer marking	other curricular			student judgements about their performance informed by parental interpretations, evaluation data
Uselman (1996)	Low	High	Portfolio, self or peer marking	cross-curricular	achievement improved through portfolios, peer evaluation and marking		improved attitudes to learning and student relationships
Young et al. (1997)	Medium	High	Formative, portfolio	literacy			participation increased

than test scores alone and improving the student-teacher dialogue which in turn raises achievement. Goodrich's (1997, high WoE) study suggests that deep processing of information, not self regulation, is the key to learning but that this needs further research.

The use of dynamic self assessment with an emphasis on mediated interactions is a form of assessment 'fit for purpose' when what is being assessed is not performance, but rather a human capability, or a process of personal development. In subsequent research which addresses the assessment of learning dispositions, teachers did not see this as an alternative to attainment but rather as a necessary prerequisite for attainment, and as a vehicle for enabling students to become aware of their own learning identities and to take responsibility for their own learning trajectories (Deakin Crick 2007). By engaging in self assessment, students are developing a language with which to describe the processes of learning, and this is an important foundation for self awareness and ownership.

4.3 In-depth review: quality assurance results and possible weaknesses

The quality assurance procedures were as described in 2.3.5 above. All the studies in the in-depth review were checked for inclusion by two members of the Review Team. Each of the studies was then data-extracted by two members of the team and entered on to REEL, the EPPI-Centre database of educational research. Eight studies were also data-extracted by EPPI-Centre staff to ensure consistency across reviews. Any differences were resolved before an agreed version was used in the final synthesis.

4.4 Summary of results of synthesis

Pupil outcomes

Most studies reported some positive outcomes including:

- 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.

CHAPTER FIVE

Implications

5.1 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) 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.

The study design further limits the transferability of the findings of some of the studies, though in some cases, the smaller scale studies offered detailed insights into the processes involved and so made a different important contribution. 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 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 the variables that contributed to any outcomes reported in order to demonstrate the effects specifically 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, low WoE). Ezell et al. (1999, medium WoE) commented on the lack of support in the education system for portfolio work.

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 that have been published in the intervening period were not included.

5.2 Implications for policy, practice and research

5.2.1 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 incorporates 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 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. The evaluation of these broader outcomes presents a challenge.
- 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.

5.2.2 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 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 this is derived from limited evidence, it suggests a need for more dialogue between parent, teacher and student.

5.2.3 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 the to 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.

CHAPTER SIX

References

6.1 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.

6.2 Other references used in the text of the Technical Report

APA Task Force On Psychology In Education (1993) *Learner-centered psychological principles: a framework for school reform and redesign* (Revised Edition). Washington DC: American Psychological Association and Mid-Continent Regional Educational Laboratory.

APA Work Group Of The Board Of Educational Affairs (1997) *Learner-centered psychological principles: a framework for school reform and redesign* (Revised Edition). Washington DC: American Psychological Association.

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.
- Hattie J, Biggs J, Purdie N (1996) Effects of learning skills interventions on student learning: a meta analysis. *Review of Educational Research* **66**: 99-136.
- 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.
- Royal Society Of Arts (2005) *Opening minds: giving young people a better chance*. London: Royal Society of Arts.
- 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 2.1: Inclusion and exclusion criteria

Inclusion criteria

The search for 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. Both naturally occurring and researcher-manipulated evaluation study types were considered to be relevant. Designs included comparison of the experience of comparable classes with different experiences of self or peer assessment or comparison of the same groups before and after the introduction of self or peer assessment. Surveys of students' and teachers' perceptions of the impact of student self or peer assessment and case studies reporting experiences and impacts of involving students in assessing their work were also considered relevant.

Exclusion criteria

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).
- C: Not reporting impact on students of the process of self- or peer assessment but just the outcome of the assessment.
- 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.

Appendix 2.2: Search strategy for electronic databases

The search of bibliographic databases used a combination of the following terms.

Self assessment and peer assessment	Relevance to school	Impact on student (dependent variable)
Student self assessment	School	Achievement
Student self evaluation	Infant school	Attainment
Peer assessment	Foundation stage	Learning outcomes
Peer evaluation	Primary School	Study skills
Learner-centred	Elementary school	Motivation
Student involved	Secondary school	Learning style
	Community school	Responsibility
	Urban school	Self-directed learning
	Suburban school	Metacognition
	Private school	Learning power
	State school	Emotional literacy
	High school	Disposition
	Middle school	Values
	Pre-school	Attitudes
	Kindergarten	Personal development

The search strategy is set out below

Database	Searching strategies	Hits
AEI	student & self assessment student & self evaluation student & peer assessment student & peer evaluation	541
BEI	student & self evaluation (sole: 889) = 160	
	student & self assessment (sole: 206)= 40	
	student & peer evaluation (sole: 235) = 99 A	
	student & peer assessment (sole: 76) = 45	
		240
ERIC (1990- Sept. 2004)		
	If AEI/BEI strategies were used, student & self evaluation = 2792	2792
	student & self assessment = 788	788
	student & peer evaluation = 582 A	582
	student & peer assessment = 103	103
	“OR”	
	AND “school” B	1394
	And “attitude” = 51	
	and “motivation” = 135	
	and “values”/”value” = 72	
	and “responsibility” = 115	
	and “disposition” = 2 (if “disposition” as the only key term = 359)	
	and “personal development” = 6	
	and “achievement” = 321	
	and “attainment” = 19 C	
	and “learning outcomes” = 17	
	and “study skills” = 27	
	and “learning style” = 10	
	and “self directed learning” = 10 (when hyphen was used = nil)	
	and “meta-cognition” = nil; “metacognition” = 40	
	and “learning power” = nil (if “learning power” as the only key term = 13)	
	and “emotional literacy” = nil (if “emotional literacy” as the only key term = 33)	
	Combine these three categories/criteria in OR =	585
ERIC (1984-1989)	same strategies as ERIC (1990-2004)	
	A	1237
	B	428
	C	195
ERIC (1966-1983)		
	If AEI/BEI strategies were used, student & self evaluation	
	student & self assessment	
	student & peer evaluation A	
	student & peer assessment	2556
	AND YR=1980/1981/1982/1983 B	674

	AND "SCHOOL" = 819	C	220
	And "attitude"		
	and "motivation"		
	and "values"/"value"		
	and "responsibility"		
	and "disposition"		
	and "personal development"		
	and "achievement"		
	and "attainment"	D	
	and "learning outcomes"		
	and "study skills"		
	and "learning style"		
	and "self directed learning"		
	and "meta-cognition"		
	and "learning power"		
	and "emotional literacy"		
	Combine these FOUR categories/criteria in OR =		77
MLA	If AEI/BEI strategies were used, student & self evaluation (term anywhere)		670
PsycINFO	student & self assessment		233
	student & peer evaluation	A	31
	student & peer assessment		166
CRESST	self evaluation/assessment; peer evaluation/assessment		nil
WoK (SSCI)	If AEI/BEI strategies were used, student & self evaluation (term anywhere)		50
1980-2005Jul.	student & self assessment		87
	student & peer evaluation	A	8
	student & peer assessment		32
ZETOC	direct import		
	If AEI/BEI strategies were used, student & self evaluation (term anywhere)		92
	student & self assessment		181
	student & peer evaluation	A	75
	student & peer assessment		79
	Total: 6829 (first group of strategies)		
second group of strategies (after consultation with ALRSG)			
	(1) using verbs: self/peer evaluate/assess in ERIC/BEI/AEI		
	(2) using verbs: self/peer evaluat*/assess* in MLA/PsycInfo (where wild cards are allowed)		
	(3) using more key words: self/peer rating/markings/grading/scoring		
	Total: 94		
Other Database searching CRESST, NFER, Scottish educational research			
	Results: CRESST - 1 selected (JS)		

Appendix 2.3: Journals handsearched

Handsearching

American Educational Research Journal (1980-1999)
American Journal of Education (1980-2001)
Art Education (1980-1999)
British Journal of Educational Studies (1980-1999)
British Journal of Sociology of Education (1980-2001)
Child Development (1980-1999)
Comparative Education (1980-2001)
Comparative Education Review (1980-2001)
Curriculum Inquiry (1980-1999)
Education Evaluation and policy analysis (1980-1999)
Educational Researcher (1980-1999)
Elementary School Journal (1980-1999)
Journal for Research in Mathematics Education (1980-1999)
Journal of Negro Education (1980-2001)
Oxford Review of Education (1980-2001)
Reading Research Quarterly (1980-1999)
Review of Educational research (1980-1999)
Review of Research in Education (1980-1999)
Studies in Art Education (1980-1999)

Keywords: self assessment/evaluation
peer assessment/evaluation

Screening procedures

(1) delete duplicates

(2) searching functions within EndNote: delete any pre-1980 reference
career development
higher education
student teacher
teacher education
nursing education
vocational education
nurse education
dental students
teacher evaluation
faculty evaluation
engineering

(3) RefViz

(4) reading abstracts and other info. GYU, RDC (inter-scoring)

APPENDIX 2.4 EPPI-Centre keyword sheet, including review-specific keywords

V0.9.5

A1. Identification of report

Citation
 Contact
 Handsearch
 Unknown
 Electronic database (please specify)

A2. Status

Published
 In press
 Unpublished

A3. Linked reports

Is this report linked to one or more other reports in such a way that they also report the same study?
 Not linked
 Linked (please provide bibliographical details and/or unique identifier)

A4. Language (please specify)

.....

A5. In which country/countries was the study carried out? (please specify)

.....

A6. What is/are the topic focus/foci of the study?

Assessment
 Classroom management
 Curriculum*
 Equal opportunities
 Methodology
 Organisation and management
 Policy
 Teacher careers
 Teaching and learning
 Other (please specify)

A6a. Curriculum

Art
 Business studies
 Citizenship
 Cross-curricular
 Design and technology
 Environment
 General
 Geography
 Hidden
 History
 ICT
 Literacy - first language
 Literacy further languages
 Literature
 Maths
 Music
 PSE
 Physical education
 Religious education
 Science
 Vocational
 Other (please specify)

A7. Programme name (please specify)

.....

A8. What is/are the population focus/foci of the study?

Learners
 Senior management
 Teaching staff
 Non-teaching staff
 Other education practitioners
 Government
 Local education authority officers
 Parents
 Governors
 Other (please specify)

A8a. Age of learners (years)

0-4
 5-10
 11-16
 17-20
 21 and over

A8b. Sex of learners

Female only
 Male only
 Mixed sex

A9. What is/are the educational setting(s) of the study?

Community centre
 Correctional institution
 Government department
 Higher education institution
 Home
 Independent school
 Local education authority
 Nursery school
 Post-compulsory education institution
 Primary school
 Pupil referral unit
 Residential school
 Secondary school
 Special needs school
 Workplace
 Other educational setting (please specify)

A10. Which type(s) of study does this report describe?

A. Description
 B. Exploration of relationships
 C. Evaluation
 a. naturally-occurring
 b. researcher-manipulated
 D. Development of methodology
 E. Review
 a. Systematic review
 b. Other review

Review specific keywords: What is the impact of self and peer assessment on student outcomes?

B1. Subject context of assessment

- English
- Mathematics
- Science
- Design and technology
- ICT?
- History
- Geography,
- Art and design
- Music
- Physical education
- Modern foreign language
- Citizenship
-

B2. Process focus of the assessment

- Acquisition of knowledge
- Learning to learn
- Metacognition
- Acquisition of skills
- Development of understanding
- Team building - interpersonal skills
- Values development
- Other
-

B3. Types of assessment

- Summative
- Formative
- Record of achievement
- Profile
- Portfolio
- Personal report
- Journal
- Other
-
-
-
-

B4. Level of ownership of the assessment process

- Low (no genuine ownership by the students)
- Medium (students adopt the goals and criteria identified by the teacher)
- High (students have ownership of the goals and are committed to engaging in learning to achieve them)
-

B5. Type of study outcome (impact on students)

- Achievement
- Enjoyment (motivation, interest)
- Engagement in learning (responsibility, commitment, self-direction)
- Social engagement
- Confidence to participate in learning community
- Well being (emotional, mental, physical)
- Other
-

Appendix 3.1: Details of studies included in the systematic map

The table below indicates the source of the initial papers found and, for comparison, the source of the studies that were included in keywording.

Results of initial search (1236 papers)

Identification	Number of hits from initial searches	Number included in keywording
Two-stage screening		
AEI and BEI	240	7
ERIC (1990-Sept. 2004)	585	15
ERIC (1984-89)	195	7
ERIC (1980-83)	77	3
MLA and PsycINFO	206	6
SSCI of ISI Web of Knowledge	177	9
British Library ZETOC	217	3
One-stage screening		
Specialist websites (NFER, CRESST, SCRE)	0	0
Handsearch (not JOL)	4	0
Journals online (JOL) in IngentaConnect and JSTOR	1	0
Personal contacts	2	1
Total*	1,236	51

*Notes:

(1) The numbers in the first column were not mutually exclusive to each other; in other words, there were duplicates amongst different databases and methods of identification (database or handsearching). All items found through handsearching of hard copies and electronic full texts online were already included in the database search. The duplicates were expunged first through EndNote's built-in function and then by visual examination of each reference, without taking into consideration which database or method they were originally identified. Therefore, the duplicates were randomly expunged, and the numbers in the second column did not necessarily indicate which database or method had more studies located, in absolute terms.

(2) No search can produce only what we want. There are always some references that are irrelevant to the review question, no matter how precise the search terms are set up. For example, using "student" and "self evaluation" also produced references about the role that "student" played in staff or faculty "self evaluation", or "student" teacher's "self-evaluation" in initial teacher training. Therefore, we developed a pre-screening strategy by reading the full bibliographic details of all the 1236 references found from the databases, SSCI and ZETOC in order to make sure that only references that were about students' self and peer assessment or evaluation in a broad sense were subject to the application of inclusion and exclusion criteria. As a result, 214 references were identified for the next stage screening process (see Figure 3.1)

Countries in which the studies took place

Country of study	Number
US	32
Canada	5
England*	3
Australia*	2
Latvia	2
Portugal	2
Barbados	1
Finland	1
Puerto Rico	1
Scotland	1
South Africa	1

Note: * One study focused on both Australia and England

Classification of the subject focus of assessment described in the study (categories not mutually exclusive)

Attribute	Number
English	10
Mathematics	13
Science	3
ICT	1
History	1
Geography	1
Art and design	1
Music	2
Physical education	2
Modern foreign language (please state)	2
Other (please state)	4
Literacy	3
Not subject-specific	11
Other	1

Appendix 4.1: Details of studies included in the in-depth review

Bickmore DK (1981) The effects of student self-evaluation and pupil-teacher conferences on student perceptions, self concepts and learning

Study type

Exploration of relationships

Effects of student self-evaluation and pupil-teacher conferences on the student perceptions, self-concepts and learning

Evaluation: researcher-manipulated

Subject context of assessment

Modern foreign language (Spanish)

Focus of the assessment

Acquisition of knowledge (language skills)

Metacognition

Self confidence

personal feelings

Types of assessment

Self or peer marking/grading categories (though not grades)

Level of ownership

High

Type of study outcome

Achievement

Purpose of the assessment?

Formative

Aims of the study?

To develop, design and implement instruments to

evaluate the effects of student self-evaluations and pupil-teacher conferences on student perceptions, self-concepts and learning

Research questions and/or hypotheses?

Explicitly stated

Seven questions (although not all relevant to this particular review question)

1. Were there differences in the amount of increase in knowledge of subject matter between the experimental and control classes at the end of the semester?

2. Were there differences in the amount of change in student self-concepts between the experimental and control classes at the end of the semester?

3. Were there differences in the amount of change in the teacher perception of student capabilities between the experimental and control classes at the end of the semester?

4. Were there differences in the amount of change in the student perception or teacher perception of student capabilities between the experimental and control classes at the end of the semester?

5. Did the students prefer this process of student self-evaluations and pupil-teacher conferences to other systems of learner evaluation they had previously operated under?

6. Did the teachers involved in the study prefer this evaluation procedure of self-evaluations and pupil-teacher conference to procedures they had formerly used?

7. What did the instructors and pupils see as the major strengths and weaknesses to this evaluation procedure of student self-evaluations and pupil-teacher conferences?

Variables or concepts measured or examined?

Explicitly stated

1. Students' knowledge increase/change
2. Students' self-concepts (e.g. of teacher perception of student capabilities).

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

Medium

Weight of evidence D:

Medium

Brookhart SM (2001) Successful students' formative and summative uses of assessment information

Study type

Evaluation: Researcher-manipulated. A rigorous study exploring how particular assessment events were perceived by students, and the way students use assessment information.

Subject context of assessment

English and Science

Focus of the assessment

Metacognition

Understanding of knowledge

Types of assessment

Students' perceptions of different *Types of assessment* that were already taking place

Level of ownership

Low

Type of study outcome

Achievement

Enjoyment

Engagement in learning

Purpose of the assessment?

Formative and summative

Aims of the study?

To record students' perceptions of assessment, both formative and summative. The authors emphasise these are successful students.

Research questions and/or hypotheses?

Explicitly stated

How do successful students perceive assessments?

Variables or concepts measured or examined?

Students' perceptions of different kinds of assessment and, implicitly, the effects of self assessment on learning

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

High

Weight of evidence D:

Medium

Bruce LB (2001) Student self-assessment: encouraging active engagement in learning

Study type

Exploration of relationships

Evaluation: researcher-manipulated

Subject context of assessment

Not subject specific, although it is subject related. The focus is more on the methods of self assessment than on a particular subject(s): there were five elective subjects reported in the dissertation.

Focus of the assessment

Learning to learn

Ownership and engagement of learning

Metacognition

Self awareness

Types of assessment

Formative. Other training and use of self assessment, and its impact on the ownership and engagement of learning.

Level of ownership

Medium

Type of study outcome

Engagement in learning

Purpose of the assessment?

Formative

Aims of the study?

Explicitly stated

In the light of the tendency of many secondary students to display passive or pragmatic approaches to formal learning situations, this study proposed the use of student self-assessment activities to elicit more student engagement in learning.

Research questions and/or hypotheses?

General hypothesis:

The introduction of student self-assessment activities - criteria-oriented and reflection-oriented ones as defined in this study - in secondary classrooms that promote relevance in coursework and an atmosphere of respect will result in students becoming more actively engaged in their learning, taking more personal ownership, and displaying less passivity or pragmatism in learning situations.

Research Questions:

Primary research questions:

1. Can the use of self assessment encourage secondary level students to become more engaged in their learning, take responsibility for their performance, and experience a genuine interest in the subject matter?
2. Can teachers learn how to use student self-assessment effectively in their classes?

Secondary questions:

1. How do students perceive this experience?
2. What kinds of challenges for teachers arise in the implementation of student self assessment?

Variables or concepts measured or examined?

1. Qualitative data

- Teachers recorded their observations and reflections related to each implementation of a student self-assessment activity - evidence of increased ownership of learning; increased enthusiasm for the subject matter; student-initiated effort to improve work; and the relating of the subject matter with other areas.
- Summative reflections from the teachers were requested at the conclusion of the study.
- Course evaluation survey given to students seeking their perception about the classroom environment and the helpfulness of student self-assessment.

- Interviews of a sample of students by the researcher to clarify patterns that emerged in early data analysis.
- Summaries of the researcher's observations from the ongoing coaching sessions with the teachers and classroom visits - evidence of students taking ownership of learning, as well as any comments made by students

2. Quantitative data

- student course survey - demographic questions and a Likert-scale evaluation of perceptions of the classroom atmosphere.
- A quantitative measure of learner attitudes sought through a pre- and post-test use of a commercial instrument. The INCLASS Inventory of Classroom Style and Skills (Miles and Grummon 1999) assessed interest in learning, having a sense of quality, taking responsibility, persisting, working in teams, solving problems, and adapting to change. The first four of these characteristics defined the qualities that made up the focus of this intervention.

Weight of evidence A:

High trustworthiness

Weight of evidence B:

High

Weight of evidence C:

High

Weight of evidence D:

High

Carter CR (1997) Assessment: shifting the responsibility*Study type*

Evaluation: researcher-manipulated

Subject context of assessment

Mathematics

Focus of the assessment

Acquisition of knowledge

Metacognition

Acquisition of skills

Types of assessment

Formative

Summative

Self or peer marking/grading

Level of ownership

High

Type of study outcome

Achievement

Pupils' understanding of progression within the subject

Engagement in learning

Purpose of the assessment?

Formative and summative

Aims of the study?

Students perform a written analysis of their corrected tests, especially errors, prior to receiving a final grade. Improvements in student attitudes toward tests and performance on them are reported from using 'Test Analysis' with gifted high school students in a calculus class.

Research questions and/or hypotheses?

To study the process of using the Test Analysis method and the impact on students and teacher

Variables or concepts measured or examined?

Student written responses given in the test analysis procedure; teachers' actions in conducting the procedure; students observed reactions to the procedure. The study is claiming that by involving the students in analysing their own tests they can become more effective learners. It addresses three areas which directly affect the gifted learner: critical analysis and critical thinking skills, the autonomous learner, and delayed gratification.

Weight of evidence A:

Low trustworthiness

Weight of evidence B:

Low

Weight of evidence C:

Low

Weight of evidence D:

Low

Crouch DW, Ward P, Patrick CA (1997) The effects of peer-mediated accountability on task accomplishment during volleyball drills in elementary physical education

Study type

Evaluation: Researcher-manipulated

Subject context of assessment

Physical education

Focus of the assessment

Acquisition of skills

Types of assessment

Self or peer marking/grading

Level of ownership

Medium

Type of study outcome

Achievement

Purpose of the assessment?

Formative

Aims of the study?

To assess the effects of group instruction, peer dyads and peer-mediated accountability on performance in specific volleyball skills

Research questions and/or hypotheses?

What is the relationship between task and consequences created by the peer-mediated accountability variable?

Variables or concepts measured or examined?

(a) the number of trials performed, and (b) the number of those trials that were correct, during one-minute trials of the forearm pass.

Three different 'conditions' were tested to see the effects on performance: group instruction, this is where students had to progress through a 'circuit of stations' completing the task; peer-dyads, where students completed the task in pairs; and peer-mediated accountability, where students kept a record of their partner's performance of the task. Researches the effects of peer assessment on performance.

Weight of evidence A:

Medium

Weight of evidence B:

Medium

Weight of evidence C:

Medium

Weight of evidence D:

Medium

Deakin Crick R, Broadfoot P, Claxton G, Yu G (unpublished) Developing 'learning power' in the classroom: the Effective Lifelong Learning Inventory (ELLI) in practice

Study type

Description

Exploration of relationships

Evaluation: researcher-manipulated

Subject context of assessment

Not subject-specific

Measuring and nurturing learning power of the students

Focus of the assessment

Learning to learn

Types of assessment

Formative self or peer marking/grading

Self grading on an ELLI questionnaire

Level of ownership

High

Type of study outcome

Engagement in learning; learning to learn

Purpose of the assessment?

Formative

Aims of the study?

The aim of this study was to explore how teachers used the self-report assessment data generated by the ELLI Learning Profile to interact with individuals and groups of learners in meeting their learning needs.

Research questions and/or hypotheses?

How do teachers respond to student self-assessment data about their own learning power, and are those teachers' interventions successful?

Variables or concepts measured or examined?

Students' self report of their learning power on seven dimensions: changing and learning, meaning making, critical curiosity, creativity, learning relationships, strategic awareness and resilience

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

High

Weight of evidence C:

Medium

Weight of evidence D:

Medium

Ezell D, Klein CE, Ezell-Powell S (1999) Empowering students with mental retardation through portfolio assessment: a tool for fostering Self-determination skills

Study type

Evaluation: naturally occurring

Portfolios are already part of the assessment array but their use was evaluated in this study

Subject context of assessment

Cross-curricular

Focus of the assessment

Learning

Types of assessment

Portfolio

Level of ownership

Low

Type of study outcome

Engagement in learning; Learning to Learn

Purpose of the assessment?

Use of portfolios to foster self determination

Aims of the study?

To investigate whether portfolio assessment can be used to foster self-determination

Research questions and/or hypotheses?

Does the portfolio assessment process contribute to self-determination in people with learning disabilities? ('Mental retardation' is the phrase used by the authors.)

Variables or concepts examined?

Self-determination skills in relation to the portfolio assessment process

Teachers, parents and learners' views of self determination before, during and after use of portfolios

Weight of evidence A:

Medium

Weight of evidence B:

High

Weight of evidence C:

Medium

Weight of evidence D:

Medium

Goodrich HW (1997) Student self-assessment: at the intersection of metacognition and authentic assessment

Study type

Exploration of relationships

Relevant to review:

RQ3: Does self-assessment influence metacognitive engagement in the classification talk?

RQ4: Does self-assessment influence learning about classification and arthropods?

Evaluation: researcher-manipulated; two groups: exp. and control

Subject context of assessment

Science classification system (e.g. arthropods)

Focus of the assessment

Acquisition of knowledge i.e. classification system;

Metacognition planning, self-check/correction, process goal-setting, perceptions of the task difficulty, demands and nature of the tasks.

Types of assessment

Formative

Self or peer marking/grading using scoring rubrics - four levels in three categories

Level of ownership

Medium

Type of study outcome

Achievement effects of self assessment on learning outcome

Metacognition in particular, of self-assessment activities

Purpose of the assessment?

Formative

Aims of the study?

To build a model of self-assessment based on contemporary educational and psychological literatures

To investigate the validity of that model through clinical research

(The author began by formulating a model of how self-assessment functions in learning, then she presented and discussed the results of the study of the effects of self-assessment on content learning and metacognition).

Research questions and/or hypotheses?

This study tests the hypothesis that task-specific, criterion-referenced self-assessment will increase metacognitive engagement and learning. This hypothesis is based on a model of self-assessment developed in this paper which states that self-assessment serves learning by increasing metacognition and deep processing.

The study focused on four research questions:

1. Do students spontaneously self-assess when engaged in a classification task?
2. What kinds of self-assessment are students capable of under supportive conditions?
3. Does self-assessment influence metacognitive engagement in the classification task?
4. Does self-assessment influence learning about classification and arthropods?

Variables or concepts measured or examined?

Before and after the intervention, instruments were used to measure students' content knowledge, metacognitive engagement, and self-assessments included: pre- and post-tests, transcriptions of think aloud protocols, and written notes on their final classification systems.

Weight of evidence A:

High trustworthiness

Weight of evidence B:

High

Weight of evidence C:

High

Weight of evidence D:

High

Gregait LH, Johnsen DR, Nielsen PS (1997) Improving evaluation of student participation in physical education through self-assessment

Study type

Evaluation: researcher-manipulated

Subject context of assessment

Physical education

Focus of the assessment

Student self-evaluation of their performance and participation

Types of assessment

Journal

Self or peer written comments

Level of ownership

Low

Type of study outcome

Participation on self-assessment

Purpose of the assessment?

Formative

Aims of the study?

To study the impact on students' ability to self-assess as a result of treatment related to self assessment in physical education - co-operative group strategies, use of journals, goal setting and participation checklists.

Research questions and/or hypotheses?

As a result of student self-assessment during the period of September 1996 to January 1997, the target groups will improve the ability to self assess participation in physical education, as measured by journals, participation checklist/rubric and surveys.

Variables or concepts measured or examined?

Students' self-assessment of participation.

Weight of evidence A:

Low trustworthiness

Weight of evidence B:

Low

Weight of evidence C:

Low

Weight of evidence D:

Low

Hewitt MP (2002). Self-evaluation tendencies of junior high instrumentalists

Study type

Evaluation: Researcher-manipulated

Subject context of assessment

Music

Focus of the assessment

Acquisition of skills

Self-assessment accuracy

Types of assessment

Self or peer marking/grading

Students' ratings were compared with adjudicators' ratings

Level of ownership

Low

Type of study outcome

Achievement

Ability to self-assess accurately

Purpose of the assessment?

Formative

Aims of the study?

(a) to determine the nature of junior high school instrumental music students' self-evaluation tendencies over time

(b) to examine whether the process of self-evaluation, with or without the use of a model, has an effect on self-evaluation accuracy

(c) to determine if a relationship exists between self-evaluation and music performance achievement

Research questions and/or hypotheses?

How do students self-evaluations change over time?

Do students' self-evaluation, with or without models, become more accurate through the process of self-evaluation?

Is there a relationship between self-evolution accuracy and music performance achievement?

Variables or concepts measured or examined?

Music performance, self-evaluation. The Woodwind Brass Solo Evaluation Form (WBSEF) was used by adjudicators to judge the students' individual performances. An adapted version of this was used by students for self-evaluation. Both forms used a criterion referenced scale for seven aspects: tone, intonation, technique, melodic accuracy, rhythmic accuracy, tempo and interpretation.

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

Medium

Weight of evidence D:

Medium

Katstra J, Tollefson N, Gilbert E (1987)
The effects of peer evaluation on attitude toward writing and writing fluency of ninth grade students

Study type

Evaluation: researcher-manipulated

Subject context of assessment

English

Focus of the assessment

Acquisition of skills

Attitude to writing

Types of assessment

Formative

Self or peer oral comments

Level of ownership

Medium

Peer-assessment modelled by the teacher

Type of study outcome

Achievement fluency in writing

Enjoyment

Attitude to writing

Purpose of the assessment?

Formative

Aims of the study?

To determine whether peer evaluation could improve attitudes toward writing and increase the writing fluency of ninth grade students.

Research questions and/or hypotheses?

It was hypothesised that there would be a significant difference between ninth grade students who peer evaluate and students who do not peer evaluate on a measure of attitude towards writing and a measure of writing fluency.

Variables or concepts measured or examined?

Attitudes to writing as measured by two instruments and fluency of writing as indicated by word count.

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

High

Weight of evidence C:

High

Weight of evidence D:

High

Kitsantas A, Reiser RA, Doster J (2004)
Developing self-regulated learners: goal setting, self-evaluation, and organizational signals during acquisition of procedural skills

Study type

Evaluation: researcher-manipulated

Subject context of assessment

ICT

Focus of the assessment

Acquisition of knowledge

Learning to learn

Metacognition

Types of assessment

Formative

Self or peer marking/grading

Level of ownership

Medium

Type of study outcome

Achievement

Engagement in learning

Purpose of the assessment?

Formative

Aims of the study?

The purpose was 'to examine the effects of goal setting (process goals vs. outcome goals), self-evaluation (presence vs. absence) and organisational signals (presence vs. absence) on students' acquisition of a procedural skills (the ability to animate slides created via presentation software), as well as the effects of these variable on students' self-efficacy, satisfaction with their performance, rating of the instruction they received, and the attribution of success or failure in acquiring the skills they were taught.' (p272)

Research questions and/or hypotheses?

That 'students assigned to the process goal condition would demonstrate a higher level of animation skills, report greater self-efficacy beliefs and satisfaction with their performance, attribute their success to strategy use and rate the quality of the instructional module more positively than those students assigned to the outcome goal condition.

We also expected that the presence of self-evaluation and organisational signals would enhance students' animation skills and affective outcomes.

Finally, regarding student attributions, we hypothesised that participants assigned to the process goal condition would attribute their success to strategy effectiveness rather than to ability, practice or effort.' (p272)

Variables or concepts measured or examined?

Animation skills

Self-efficacy satisfaction with their performance rating of the instruction received

Attribution of success or failure in acquiring the skills

Weight of evidence A:

High trustworthiness

Weight of evidence B:

High

Weight of evidence C:

Medium

The operational meaning of self-evaluation is not in accordance with the idea of students making their own judgements of their progress. In this study they merely gain access to feedback on whether or not they used the skill correctly.

Weight of evidence D:

Medium

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

Study type

Evaluation: researcher-manipulated

Controlled trial with 133 students, allocated to six treatment groups. Pre- and post-intervention tests.

Subject context of assessment

English

Science

History

Geography

Focus of the assessment

Learning to learn

Metacognition

Development of understanding

Types of assessment

Formative

Self or peer written comments

Level of ownership

High

Type of study outcome

Achievement

Engagement in learning

Confidence to participate in learning community

Purpose of the assessment?

Formative

Aims of the study?

To investigate the effects of exposure to a concept in multiple subject areas and metacognition self-monitoring training on the transfer of middle school students

Research questions and/or hypotheses?

(a) students who possessed good schemata would perform better on the transfer tasks than would students with poor schemata

(b) students who were exposed to two subject areas and were trained in self-monitoring would have better schemata than would both one-subject self-monitoring-trained students and two-subject not self-monitoring trained students

(c) students who reported engaging during the learning and transfer tasks in high levels of metacognitive activity would perform better on the transfer task than would other students

(d) two-subject self-monitoring students would self-monitor better (i.e. report higher levels of metacognitive activity) than would one-subject self-monitoring students (as well as all other students)

(e) two-subject self-monitoring students would outperform all other students on the transfer tasks

Variables or concepts measured or examined?

Transfer ability and schema (i.e. the understanding of the nature and use of concept mapping)

Weight of evidence A:

Medium trustworthiness

This study is very carefully executed and meticulously reported. However, it adopts very limited views of what self-monitoring is and elevates concept mapping to something that signifies higher-

level skills. In other words, both treatment and outcome are somewhat inconsequential.

Weight of evidence B:

Medium

Overall the design might be appropriate, but the length and nature of the treatment is limited.

Weight of evidence C:

Low

The way in which the variables are operationalised leave room for doubt that the study really addresses the review question.

Weight of evidence D:

Low

Klenowski V (1995) Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England

Study type

Evaluation: naturally occurring

Subject context of assessment

The study included students and teachers from a range of subjects within the science department in the English FE college and teachers from a range of subjects in the Australian secondary school.

Focus of the assessment

Acquisition of knowledge

Learning to learn

Metacognition

Acquisition of skills

Interpersonal skills

Values development

Types of assessment

Formative

Portfolio

Level of ownership

High

Type of study outcome

Achievement

Engagement in learning

Social engagement

Confidence to participate in learning community

Purpose of the assessment?

Formative

Aims of the study?

'To discover the potential of alternative forms of formative evaluation, such as student self-evaluation' (p149)

Research questions and/or hypotheses?

'The study was exploratory, aimed at generating some understanding of the complexity of student self-evaluative processes' (149)

'This research is based on the assumption that students' commitment to learning is likely to be strengthened when they take more responsibility, in collaboration with their teachers, for monitoring their own progress, for evaluating their own strengths and weaknesses, and for collectively devising strategies for improvement of the learning outcomes.' (p146)

Variables or concepts measured or examined?

Impact on pedagogy

Student-teacher relationships

Student responsibility for their learning

Student metacognition

Student ability to use criteria in self and peer assessment

Student self-esteem

Students' self-awareness, confidence and understanding of their strengths and areas for improvement

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

High

Weight of evidence D:

Medium

Knubb-Manninen G (1994) Processing habits and second language learning: students' self-evaluation of their learning

Study type

Exploration of relationships

A cross sectional case study of five different groups of learners, differentiated by success in school and success in second language learning.

Subject context of assessment

Modern foreign language

Focus of the assessment

Learning to learn

Metacognition

Development of understanding

Types of assessment

Formative

Level of ownership

High

Type of study outcome

Achievement

Purpose of the assessment?

Formative

Aims of the study?

To analyse students' conscious second language learning activity and the kinds of language problems the students themselves recognised

Research questions and/or hypotheses?

How self-evaluated processing problems are related to learning contexts, learning habits and learning success

Variables or concepts measured or examined?

Students' conceptions of their own learning difficulties in language learning and use

Weight of evidence A:

High trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

Low

There is little direct evidence of the effect of self-evaluation - the study is about the results of self-evaluation. Its relevance is in the interpretation that the students were not actually actively self-evaluating.

Weight of evidence D:

Low

Marshall J (1993) Self-assessment in secondary schools: a study

Study type

Evaluation: researcher-manipulated

Subject context of assessment

English

First language

Focus of the assessment

Acquisition of skills

Literacy (English class: mainly in relation to assignment tasks)

The impact of self-assessment were reported as: motivation, confidence, etc.

Types of assessment

Formative

Checklists of skills related to the assignment task

Completion of charts to indicate understanding of the task

Discussion with a peer followed by recording of the outcomes of that discussion

Self-reflective writing on the processes and skills used by the student in completing the set assessment task

Self-evaluation checklists

Self or peer written comments

Self or peer oral comments

Level of ownership

High

Type of study outcome

Achievement

Engagement in learning

Confidence to participate in learning community

Purpose of the assessment?

Formative

Aims of the study?

This paper explores the topic of self-assessment and its impact on a class of Year 11 English students in an all girls' school. The project was proposed as a means for finding ways to incorporate self assessment and metacognitive strategies into the learning experiences of secondary students.

Research questions and/or hypotheses?

1. Students' responses to self-assessment activities
2. The project was proposed as a means for finding ways to incorporate self-assessment and metacognitive strategies into the learning experiences of secondary students.

Variables or concepts measured or examined?

Not all these data are reported in the paper.

1. Perceptions of the students on self-assessment: after submitting their assignment to the teacher, but before receiving their teacher-assigned result, the students were asked to respond to an evaluation questionnaire. Questions included: how difficult to do did you find these activities (i.e. the six types of self-assessment activities), how useful did you find these activities, how effective were these activities in helping to improve your performance in this task, what was the effect of these activities in regard to your motivation and self-confidence in completing the task?
2. Interviews with 3 selected students
3. Interviews with the class teacher, as a means of triangulation for the purposes of cross-checking data gained from the individual students

Weight of evidence A:

High trustworthiness

Weight of evidence B:

High

Weight of evidence C:

High

Weight of evidence D:

High

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

Study type

Exploration of relationships - effects of self-assessment training on external examinations

Evaluation: researcher-manipulated

Subject context of assessment

Across subjects. The curricular areas represented were: Business Studies, Humanities , Science and Technical Studies.

Focus of the assessment

Acquisition of knowledge: results of external national examination

Types of assessment

Summative

National examination

Training of self-assessment skills and its impact on external examinations

Level of ownership

High

Type of study outcome

Achievement

Enjoyment

Engagement in learning

Purpose of the assessment?

Formative and summative: self-assessment training involves formative assessment and the external examinations involve summative measures across subjects

Aims of the study?

To examine the effects of formal self-assessment training on student performance in external examinations

Research questions and/or hypotheses?

What is the effect of formal self-assessment training on student performance in external examinations?

Variables or concepts measured or examined?

Performance in external examinations

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

High

Weight of evidence C:

High

Weight of evidence D:

High

Olina Z; Sullivan HJ; (2002) *Effects of Classroom Evaluation Strategies on Student Achievement and Attitudes*

Study type

Exploration of relationships

The effects of the three treatment conditions within a 12-lesson instructional programme on how to conduct experiments and produce research reports

Evaluation: Researcher-manipulated: three treatment conditions: (1) no evaluation (2) teacher evaluation (3) self evaluation plus teacher evaluation

Subject context of assessment

Not subject-specific skills to conduct psychological experiments and produce research reports

Focus of the assessment

Acquisition of skills

Research skills (conducting experiments and producing research reports)

Types of assessment

Formative and summative

Self or peer marking/grading

Self or peer written comments

Level of ownership

Medium

Type of study outcome

Achievement: post-test on the knowledge and skills of doing research

Enjoyment

Confidence to participate in learning community

Purpose of the assessment?

Both formative feedback from students themselves: comments on their own work and from teachers

Post-test was summative

Aims of the study?

This research aimed to investigate the effects of teacher evaluation and the combination of teacher evaluation and student self-evaluation on student performance and attitudes towards the research/ instructional programme.

Research questions and/or hypotheses?

Three research questions:

1. Does teacher evaluation have a positive effect on student performance?
2. Does the combination of teacher evaluation and student self-evaluation have a different effect on student performance than teacher evaluation alone?
3. Does the combination of teacher evaluation and student self-evaluation have a different effect on student attitudes than teacher evaluation alone?

Variables or concepts measured or examined?

Student performance in research report writing; and their attitudes towards the instructional/research programme:

1. Ratings of student research reports
2. Student post-test performance (21 multiple choice and short answer items which were directly aligned with the objectives of the instructional programme)
3. Student attitude toward the instructional/ research programme

Weight of evidence A:

High trustworthiness

Weight of evidence B:

High

Weight of evidence C:

High

Weight of evidence D:

High

Olina Z, Sullivan HJ (2004) Student self-evaluation, teacher evaluation, and learner performance

Study type

Exploration of relationships

Evaluation: researcher-manipulated

Subject context of assessment

Research skills (report writing and doing experiments)

Some studies were conducted as part of a high school psychology class

Focus of the assessment

Learning to learn

Acquisition of skills: writing research/experiment reports and doing experiments

Types of assessment

Four criterion measures were included in the study:

- (a) ratings of the student projects
- (b) post-test scores
- (c) student attitude surveys
- (d) teacher attitude surveys.

Level of ownership

Low

Type of study outcome

Achievement

Confidence to participate in learning community

Purpose of the assessment?

Formative

Aims of the study?

To build upon previous research by extending the scope of the study and also to test the ideas in a country (Latvia) with traditional teaching methods and no established procedures for student investigation or formative assessment

Research questions and/or hypotheses?

1. What are the comparative effects of teacher evaluation and student self-evaluation on student performance?
2. Does the combination of teacher evaluation and student self-evaluation improve student performance to a greater degree than either of the evaluation procedures alone?
3. Do the two evaluation procedures have differential effects on student attitudes?

Variables or concepts measured or examined?

Criterion measures include performance and perceptions (attitudes)

Four criterion measures were used in the study:

- (a) ratings of the student projects
- (b) post-test scores
- (c) student attitude surveys
- (d) teacher attitude surveys.

Weight of evidence A:

High trustworthiness

Weight of evidence B:

High

Weight of evidence C:

High

Weight of evidence D:

High

Powell SD, Makin M (1994) Enabling pupils with learning difficulties to reflect on their own thinking

Study type

Evaluation: researcher-manipulated

Subject context of assessment

Mathematics

Focus of the assessment

Metacognition

Values development

Types of assessment

Self or peer oral comments

Verbal description of their work, or significant aspects of their work, to the teacher

Level of ownership

Low

Type of study outcome

Confidence to participate in learning community.

During the last 15 minutes of the session the group would reassemble for the 'pupil self-appraisal' phase. The self-appraisals provided a formal occasion for pupils to reflect on their performance and were seen as a way of demonstrating to them their own role in learning.

Purpose of the assessment?

Formative

Aims of the study?

To study the impact of a teaching programme 'in which we gave priority to pupils' reflection on their own learning' (p2)

Research questions and/or hypotheses?

Can the performance of these pupils with Moderate Learning Difficulties be improved by introducing activities aimed at helping them to reflect on their own thinking?

Variables or concepts measured or examined?

Change in pupils' performance in tasks, awareness of their own abilities and willingness to engage with the teacher

Weight of evidence A:

Low trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

Low

Weight of evidence D:

Low

Rief L (1990) Finding the value in evaluation: self-assessment in a middle school classroom*Study type*

Description evaluation: researcher-manipulated

Subject context of assessment

English

Focus of the assessment

Acquisition of skills in reading and writing

Types of assessment

Portfolio

Level of ownership

High

Type of study outcome

Achievement

Engagement in learning

Pupils' understanding of progression within the subject

Purpose of the assessment?

Formative

Aims of the study?

To study the effect on students of giving them responsibility for selecting best writing pieces for their portfolio and select their own reading books.

Research questions and/or hypotheses?

How students respond to being given more opportunity and responsibility for self-evaluation when selecting items for their portfolio

Variables or concepts measured or examined?

The quality of students' writing and the responsibility they take for setting goals and evaluating their progress.

Weight of evidence A:

Low trustworthiness

Weight of evidence B:

Low

Weight of evidence C:

Medium

Weight of evidence D:

Low

Ross JA, Rolheiser C, Hoaboam-Gray A (1998) Impact of self-evaluation training on mathematics achievement in a cooperative learning environment*Study type*

Evaluation: researcher-manipulated

Seven intact classes were assigned to two conditions and were tested pre- and post-intervention and four weeks after intervention stopped

Subject context of assessment

Mathematics

Focus of the assessment

Acquisition of knowledge

Learning to learn

Metacognition

Acquisition of skills

Development of understanding

Types of assessment

Formative

Self or peer marking/grading

Self or peer written comments

Self or peer oral comments

Level of ownership

High

Type of study outcome

Achievement

Engagement in learning

Purpose of the assessment?

Formative

Aims of the study?

To examine the effects of student self evaluation on achievement

Research questions and/or hypotheses?

Does self-evaluation training increase the accuracy of students' self-assessments?

Does self-evaluation training contribute to mathematics achievement?

Variables or concepts measured or examined?

Achievement accuracy of self-evaluation goal orientation survey i.e. mastery, ego and affiliation

Attributions for success or failure

Student self efficacy

Weight of evidence A:
Medium trustworthiness

Weight of evidence B:
Medium

Weight of evidence C:
High

Weight of evidence D:
Medium

Ross JA, Rolheiser C, Hogaboam-Gray A (2002) Influences on student cognitions about evaluation

Study type

Evaluation: naturally occurring

Interview data were collected from students about their experiences of self-evaluation.

Subject context of assessment

Not subject-specific

Focus of the assessment

Learning to learn

Metacognition

Development of understanding

Types of assessment

Formative

Self or peer marking/grading

Self or peer written comments

Self or peer oral comments

Level of ownership

High

Type of study outcome

Engagement in learning

Social engagement

Confidence to participate in learning community

Pupils' understanding of progression within the subject

Purpose of the assessment?

Formative

Aims of the study?

This study examined children's cognition about two types of evaluation information, traditional evaluation by the teacher and self-evaluation. It focused on self-evaluation as an exemplar of alternative assessment and because they thought that students who were engaged in self-evaluation would be more articulate about how they processed evaluation data of any kind.

Research questions and/or hypotheses?

What is children's cognition about self-evaluation and teacher evaluation?

Variables or concepts measured or examined?

Students' cognition about self and teacher evaluation

Weight of evidence A:

High trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

Medium

Weight of evidence D:

Medium

Uselman J (1996) Aiding seventh and eighth graders at a private Christian school to take ownership of their own learning

Study type

Evaluation: researcher-manipulated

The researcher introduced a 'self assessment' intervention for students working across the curriculum. She then used qualitative methods to analyse the effect of this intervention on student outcomes.

Subject context of assessment

Not subject-specific

Focus of the assessment

Learning to learn

Student ownership of their own learning

Types of assessment

Portfolio

Self or peer oral comments

Assignment calendars

Student conducted conferences with parents

Level of ownership

High

Type of study outcome

Achievement

Enjoyment

Engagement in learning

Social engagement

Confidence to participate in learning community

Purpose of the assessment?

Formative

Aims of the study?

The overall aim was to teach junior high students to take ownership of their own learning.

Research questions and/or hypotheses?

The hypothesis was that giving students choice, encouraging teamwork with parents and teachers, empowering students through giving them control and responsibility would lead to increased motivation, which would result in greater ownership of their own learning.

Variables or concepts measured or examined?

Development of student ownership of their own learning through portfolio, self assessment and student led conferences. The number of students who:

(1) would decide on their grade, on the amount of time they were willing to work on assignment preparation, on discussion with the teachers and on revising assignments;

(2) kept records of their completed assignments;

(3) came to class with completed assignments on which they had spent enough time to meet their goals;

(4) would correct or revise work on which they had earned a grade of 75 percent or lower.

Weight of evidence A:

Low trustworthiness

Weight of evidence B:

Low

Weight of evidence C:

Medium

Weight of evidence D:

Low

Young JP, Mathews SR, Kietzmann AM, Westerfield T (1997) *Getting disenchanted adolescents to participate in school literacy activities: portfolio conferences*

Study type

Evaluation: researcher-manipulated - evaluation of classroom practice (in particular, portfolio assessment) through various sources of data: written records of portfolio conferences; personal preference inventories; and student interviews

Subject context of assessment

English

Focus of the assessment

Acquisition of skills: reading and writing

Types of assessment

Formative

Portfolio plus portfolio conferences

Self or peer marking/grading

Negotiation with teacher grading/comments

Level of ownership

High

Type of study outcome

Achievement

Engagement in learning

Purpose of the assessment?

Formative and summative

Aims of the study?

To describe the role the portfolio process played in an English class and to report the students' reactions to their participation in the process

Research questions and/or hypotheses?

Involvement in the portfolio process will increase motivation, learner control and engagement

Variables or concepts measured or examined?

Measuring the development of student motivation to learn

Students focus on individual goals of progress and improvement including goal setting, strategy selection, and reflection on the process of portfolio building

Sense of self-efficacy

Weight of evidence A:

Medium trustworthiness

Weight of evidence B:

Medium

Weight of evidence C:

Medium

Weight of evidence D:

Medium

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>

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