A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

Review conducted by the Assessment and Learning Research Synthesis Group

PRINCIPAL AUTHORS

This work is a review of the Assessment and Learning Research Synthesis Group (ALRSG). It was conducted following the procedures for systematic review developed by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) and in collaboration with David Gough and Dina Kiwan, and with help from other members of the EPPI-Centre education team.

Wynne Harlen, Graduate School of Education, University of Bristol
Ruth Deakin Crick, Graduate School of Education, University of Bristol

ALRSG members

Professor Paul Black, King's College, University of London
Professor Patricia Broadfoot, University of Bristol
Professor Richard Daugherty, University of Wales, Aberystwyth
Professor John Gardner, Queen's University, Belfast
Professor Wynne Harlen, University of Bristol
Dr Mary James, University of Cambridge
Dr Gordon Stobart, Institute of Education, University of London
Professor Dylan Wiliam, King's College, University of London

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

Mr P Dudley, Head of School Improvement and Lifelong Learning, Redbridge, and member of AAIA (at the time of the review)
Mr R Bevan, Deputy Headteacher, King Edward VI Grammar School, Chelmsford
Ms P Rayner, Link Adviser for Primary Education, Nottinghamshire
Dr Ruth Deakin Crick, Researcher

ADVISORY GROUP MEMBERSHIP

The ALRSG is advised by the following international experts:

Dr Steven Bakker, ETS International, The Netherlands
Dr Dennis Bartels, President, TERC, Cambridge, MA, USA
Professor Lorrie Shepard, President, AERA, 1999-2000, University of Colorado
Professor Eva Baker, co-director of CRESST, University of California, USA
Dr T Crooks, Director, EARM, University of Otago, Dunedin, New Zealand

ACKNOWLEDGEMENTS

GLOSSARY OF TERMS

Affordances	The opportunities made available (by the use of information technologies). In relation to the nature of the use of ICT for the assessment of creative and critical thinking skills, as it is reported in these studies, there were different 'affordances' of ICT which were exploited by the researchers. The term 'affordances' is used within a theoretical framework that sees ICT as a set of cultural tools offering opportunities for action (Barnes 2000, McFarlane 2001, Wertsch 1998).
Authentic rational thinking	Authentic rational thinking describes logical reasoning that is relevant to a particular situation.
Cloze procedures	Cloze procedures is used with reference to a test of readability or comprehension in which a person is required to supply words which have been deliberately omitted from a passage.
Conative	Referring to the drive or energy to do something or to strive to understand.
Concept-mapping	Concept, or knowledge, mapping involves creating links between words representing concepts in order to produce propositions about their relationship. The link is made by an arrow, which is labelled with the nature of the link. A list of concepts may be given or created by the mapper; similarly the link terms may be given or created.
Creative thinking	Involving relating together principles, ideas, information and entities in new and original ways to generate new entities or ideas.
Critical thinking	Involving the evaluation of arguments or propositions in relation to evidence, reasoning, drawing conclusions.
Formative assessment	Assessment that is part of teaching and learning and is to help learning.
Higher order thinking skills	These are mental skills of reasoning, analysis, synthesis and evaluation (Higher order thinking occurs when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations).

Metacognition	Metacognition refers to the individual's knowledge concerning his/her own cognitive processes.
Seriation	Seriation describes the action or result of arranging items in a sequence according to prescribed criteria.
Stoicheiometry	Stoicheiometry refers to the quantitative relationship between the substances in a reaction or compound.
Summative assessment	The term summative assessment refers to an assessment for the purpose of providing a record of a pupil's overall achievement in a specific area of learning at a certain time.

TABLE OF CONTENTS

SUMMARY

Background

The review reported here was prompted by the rapid changes associated with the 'information age'. New technologies have created both the need for education to provide students with what are described as 'higher level thinking skills' and the opportunity to teach and assess these skills. There is also evidence from two previous reviews of assessment (Harlen and Deakin Crick, 2002; Black and Wiliam, 1998) that, on the one hand, what is assessed for summative purposes is what is valued in the curriculum, and, on the other hand, that formative assessment of what is taught leads to improved learning. It follows that, if valued goals of education are to be taught effectively, they need to be assessed effectively for both formative and summative purposes. The reported neglect of creative and critical thinking in assessment (Harlen and Deakin Crick, 2000) is therefore a cause for concern, given the prominence it is accorded in current discussion of the education that students need in preparation for life in a rapidly changing society and for life long learning.

This review has focused on critical and creative thinking since, since despite a lack of clarity as to the meaning of higher level thinking, there is a degree of agreement that critical thinking and creative thinking are key elements of it. Both are involved in problem-solving to varying degrees depending on the nature of the problem. The definitions adopted here are as follows:

- **critical thinking:** involving the evaluation of arguments or propositions in relation to evidence, reasoning, drawing conclusions
- **creative thinking:** involving relating together principles, ideas, information and entities in new and original ways to generate new entities or ideas

The review was funded by the EPPI-Centre at the Institute of Education, University of London, and conducted on behalf of the Assessment and Learning Research Synthesis Group (ALRSG) by the authors, with the guidance of the Review Group and the participation of members at various stages.

Aims

The aims of the review were as follows:
- to gather evidence of any impact on students and teachers of the practice of using information and communication technologies (ICT) in assessing creative and critical thinking skills for both formative and summative purposes
- to determine the positive and the negative aspects of using ICT for assessing creative and critical thinking skills, in particular, any differential impact relating to student characteristics and experience
- to make recommendations for policy and practice based on these findings
- to identify questions needing to be addressed by research so that decisions on policy and practice relating to using ICT in assessment can be evidence-based.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

1

Review questions

The review was designed to answer the following main question:

- What evidence is there of the impact of the use of ICT for assessment of creative and critical thinking skills on students and teachers?

In order to achieve all the aims of the review, it was necessary to address the further subsidiary questions:

- How does any impact vary with the experience of the students and the conditions of the assessment?
- How does any impact vary according to whether the purpose of the assessment is formative or summative?
- What are the implications for assessment policy and practice of these findings?

Methods

The review was conducted according to the procedures developed by the EPPI-Centre for the review of educational research. These procedures set out well-defined stages and require careful documentation of all decisions and actions. This means that the evidence base for the findings is made explicit and the review can be updated in the future, building on and extending what has already been done.

The stages were as follows: specifying the question and conceptual framework, developing the protocol detailing methods for the review, searching for studies; applying criteria to abstracts; obtaining full texts, keywording and summarising the characteristics of studies; applying refined criteria for final selection; extracting data and evaluating weight of evidence; synthesizing and consultation. Through these stages, the studies found were filtered using various screening processes in order to identify those which provided the most sound and relevant evidence for answering the review questions.

The search for studies was guided by selection criteria. Studies were sought that were written in English, reported on the formative or summative assessment of creative and critical thinking using ICT, and concerned students aged 4 to 18 in schools. The full texts of studies potentially meeting these criteria were obtained, read, again screened against the inclusion criteria, and reasons for exclusion were recorded. Those meeting the criteria were labelled using a core set of keywords (*EPPI-Centre Educational Keywording Sheet,* see Appendix 2.3) and some review-specific keywords. This resulted in further studies being excluded, the reasons being recorded. The remaining studies were used in mapping the identified relevant research in terms of the keywords. These mapped studies were then subjected to in-depth study and analysis in which data were extracted using the *Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research* (EPPI-Centre, 2002) and review-specific questions.

Judgements were made as to the weight to be attached to the evidence from each study in relation to its methodological soundness (as reported) and to the appropriateness of its design and relevance of its focus for this review.

Quality assurance procedures involved 20 of the studies being keyworded by two people, all data-extraction being carried out independently by at least two people, and differences reconciled by discussion. The authors and four members of the Review Group took part in keywording and data-extraction. Two members from the EPPI-Centre also carried out a quality assurance role in the keywording and data-extraction processes for a sample of studies.

The synthesis of information extracted from the studies involved identifying, in the selected studies, relationships between the affordances (that is the opportunities made available by the use of the information technologies), the processes and the outcomes of using ICT for the assessment of critical and creative thinking skills. The studies provided evidence of three types of impact: impact on (i) teachers' knowledge of students' processing and learning, (ii) their summative knowledge of students' attainment, and (iii) students' achievement or performance. The main findings were reported in relation to each of these impacts. Other outcomes, inferred or used as explanatory factors by the authors of the studies, but not based on systematic evidence, are reported as implications.

Results

Identification and categorisation of studies

A total of 103 studies were found as a result of the search for studies potentially meeting the inclusion criteria. Full texts were obtained for 94 of these, from which 62 were excluded because they did not meet all of the following criteria: concerned with assessment; ICT-based; concerned with students within the age range aged 4 to 18 in school; concerned with critical or creative thinking skills; reporting research. The remaining 32 studies were keyworded, a process that resulted in a further 20 being excluded, as detailed study showed that they did not meet the inclusion criteria. Of the 12 included studies, eleven were evaluations of situations set up by the researchers, of which four involved random allocation of students to groups.

Evidence of impact

Two of the 12 studies meeting the inclusion criteria provided evidence of high weight in relation to the review question, itself a significant finding of the lack of research focusing on this question and providing dependable evidence.

Evidence from the two high-weight studies was as follows:
- Computer-based concept-mapping with automated scoring can be used to provide summative assessment of critical and creative thinking about complex relationships.
- The use of ICT can help teachers by:
    – storing and recording information about how students are developing understanding of new material;
    – taking over some of the role of assessing and providing feedback to students so that the teacher can focus on ways of supporting learning that are beyond the program's mediation.
- Feedback from the computer during the use of test material improves student performance in later use of the same test material.

The following findings were from nine studies providing medium-weight evidence in relation to the review questions and should thus be treated more cautiously. Unless otherwise indicated, the evidence is derived from individual studies:

- Use of computers to assess teamwork did not provide evidence that aspects of collaboration result in increased problem-solving (as measured by computer-based knowledge-mapping).
- Automated collection and scoring of the processes used in problem-solving provided additional information relevant to problem-solving performance.
- Using a computer program both to test and give feedback to students can increase the level of performance as compared with students taking the same tests on paper.
- Using a computer-based program for solving problems of balancing chemical equations enabled students to achieve a higher level of performance than solving equations on paper.
- A conventional multiple-choice test gave a false impression of understanding compared with the analysis of multimedia presentations on the same topic.
- Subject matter used in computer assessment of problem-solving affected the outcome for girls more than boys.
- The performance of middle-school students was not improved by training in the use of computer graphics aimed at helping them make meaning of new material.
- Use of a computer program involving diagrammatic representation provided useful information about students' causal reasoning thinking through analysis of their diagrams but not from the log files of their computer moves.
- There was conflicting evidence from two studies relating to the impact on performance of using the web to search for information.

Evidence of variation of impact relating to the conditions and purposes of testing and the experience of the students

- Several studies showed that interacting with a computer provides feedback that supports better performance even if this only reflects back to the students the moves and links they made in a visual representation of relationships. These studies provide high and medium-weight evidence.
- Computers were shown to provide information about processes in reaching a solution that gives additional feedback to students and teachers.
- There was evidence that the choice of content influences the performance of girls to a greater degree than boys.
- Working with a team at a distance was found not to improve performance in knowledge-mapping.
- Computer-based assessment enabled students to achieve at a higher level than in equivalent paper-based tasks.
- Using technology probed students' understanding to a greater degree than conventional tests.
- Students' experience with computers and attitude towards them can influence computer-based test performance.

Other impacts reported as explanatory factors of the effects of using ICT

The authors of several studies included comments on other impacts and outcomes of computer-based assessment, such as motivation, meaning-making,

awareness of learning processes and feelings of competence. Since none of the studies included quantitative evidence of these impacts or offered systematic qualitative evidence, the points made have therefore to be treated as tentative. However, they are included in the report because they point to some possible explanations of the impact of the computer on students' achievement and relate to the affordances of ICT.

- Analysis of computer concept-mapping suggests that students develop understanding incrementally, but there is a lack of evidence of students being aware of their learning processes.

- Studies showed that real-life problems can be presented and represented in ways that call upon creative and critical thinking.

- The visual representation afforded by the use of computer programs in several studies was suggested as supporting students meaning-making.

- The concrete form of abstract relationships provided by computer programs, particularly in concept-/knowledge-mapping, was considered to be a factor in improved performance.

- There were several studies which indicated that a range of products and processes can be used to assess creative and critical thinking, often through problem-solving.

- The use of ICT made practicable the collection and use of information about the moves and processes that learners use in tackling assessment tasks, thus enabling this information to be used in supporting learning.

- The use of a computer program was considered to relieve demand on working memory by providing an 'external memory' for the student.

- In several studies, authors reported information evidence of the use of computer programs resulting in improved motivation for the assessment task, either through the use of the computer *per se* or through the feedback it provided.

- The impact of human mediation may be increased by computer mediation due to the computer enhancing the feeling of competence given by the teacher mediation alone.

Implications for policy, practice and research

In order to extend the thinking in relation to implications, a short consultation conference was held with researchers, practitioners and students of ICT in education in the UK. The outcomes of this discussion are included in the following list of implications.

Policy

- The term ICT covers a range of diverse programs, applications, hardware and software, with varied affordances for assessment and for learning. It is important to recognise this and not to treat ICT as if it were a single entity.

- The use of any form of ICT in education should start from consideration of how it can be integrated into learning, teaching and assessment.
- Priority has to be given to conceptualising learning as having higher order thinking as key features as a basis for developing and assessing creative and critical thinking. Knowing how to use ICT should go hand in hand with understanding how to bring about higher level thinking.
- There is need for professional development of teachers, teacher educators and advisers in the role that ICT in its various forms can take in learning, teaching and assessing higher order thinking.
- The route to effective use of ICT for learning, teaching and assessment in the classroom is through the teachers. Thus it is important for teacher to recognise the affordances of ICT at first hand. Familiarity with using forms of ICT in their work and professional development to create and develop their own knowledge and understanding is a precursor to effective use of ICT in the classroom.
- Teachers should be provided with ways of evaluating software for its potential in developing and assessing critical and creative thinking, and opening up aspects of the subject not available elsewhere.

Practice

- Learners need to be made aware of creative and critical thinking as explicit learning objectives so that they can use the feedback provided by ICT for formative self-assessment.
- Teachers should recognise their own role in exploiting the potential of ICT to provide intrinsic motivation for learning, through facilitating self-regulation, allowing learners more control of their learning and focusing their own role on aspects of mediation that are beyond the computer.
- The role of ICT in facilitating reflection on learning processes – the development of metacognition – should be made explicit to learners.
- Teachers need to be able to review software critically in relation to its potential for providing information for formative and summative assessment of higher level thinking.
- Teachers need to be clear about the relationship/distinction between the roles that ICT can take in learning and assessment.

Research

- The small number of studies identified, with a quarter of these emanating from one institution, is an indication of the need for a greater range of research studies on impact at the school level.
- There is need also for studies with different foci, in particular on the impact of specific affordances, such as non-linear representations, or feedback to teachers and students.
- More general studies of the affordances of ICT are needed, providing qualitative data as well as quantitative data that facilitate understanding of whether and, if so, how, critical and creative thinking can be developed and assessed.
- It is important that all research is reported and not suppressed on account of lack of impact or commercial interests.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

6

1. BACKGROUND

The reasons for this review are rooted in the rapid societal changes taking place through the development of information and communications technologies (ICT). The new technologies both create the need for education in what are described as 'higher level thinking skills' and the opportunity to teach and to assess these skills. This comes at a time when policy-makers and educationists are placing more emphasis on higher order thinking skills, of which critical and creative thinking are key components. However, if more than lip-service is to be given to these cognitive skills, it is essential for them to be included in assessment that counts, that is, assessed as valued outcomes of education.

ICT can provide opportunities for students to use and develop their creative and critical thinking, for instance, by using a computer program to set problems such that alternative solutions can be created, tried and evaluated, where relevant information has to be sought and selected from irrelevant information, and the selection justified, where situations can be modelled dynamically, enabling hypotheses to be tested. In all these situations the learner's thinking can be assessed either by the teacher or another person assessing the performance or by the computer being programmed to score responses or moves automatically as feedback to the teacher and/or learner. This review includes both studies where judgements were made by the computer and where these were made by the teacher on the basis of the learner's processes and products during interaction with the technology.

This first chapter expands on the claims made for the role of technology in developing and assessing creative and critical thinking in setting out the aims and rationale of the review and in providing some discussion of the complex conceptual issues and reasons for focusing on critical and creative thinking skills. Later sections give an overview of current policy and practice and of the research findings that form the background to the review, leading to the identification of the review questions.

1.1 Aims and rationale for current review

The main aims are as follows:
- to gather evidence of any impact on students and teachers of the practice of using information and communication technologies (ICT) in assessing creative and critical thinking skills for both formative and summative purposes
- to determine the positive and the negative aspects of using ICT for assessing creative and critical thinking skills, in particular, any differential impact relating to student characteristics and experience
- to make recommendations for policy and practice based on these findings
- to identify questions needing to be addressed by research so that decisions on policy and practice relating to using ICT in assessment can be evidence-based.

This review is a natural follow-up to the ALRSG systematic review of the impact of summative assessment and testing on students' motivation for learning (Harlen and Deakin Crick, 2002) and the review of assessment and classroom learning by Black and Wiliam (1998). These two reviews point to the conclusion that, on

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

7

one hand, what is assessed summatively is what is valued in the curriculum and, on the other hand, that formative assessment of what is taught leads to improved learning. In different ways, therefore, they confirm the close connection between teaching, learning and assessment. It follows that, if valued goals of education are to be effectively taught, they need to be effectively assessed for both formative and summative purposes. The reported neglect of creative and critical thinking in assessment is therefore a cause for concern, given the prominence it is accorded in current discussion of educating students for a rapidly changing society and for lifelong learning.

In identifying the focus of this review, the ALRSG brings together three factors:
- recognition that an important aim of modern education should be to teach students to think creatively and critically, to teach them how to learn, and to think about their own learning, as a preparation for continuing learning throughout their lives;
- evidence that what is valued has to be assessed (formatively so that learning is improved and summatively as an indication of its value);
- the potential of the use of ICT to extend what can be validly assessed beyond the bounds of most current practice.

The overall purpose of the review is to find and synthesise evidence of the effect which the use of ICT to assess creative and critical thinking skills has on students and teaching.

1.2 Definitional and conceptual issues

1.2.1 Creative and critical thinking

The discussion of the development of thinking and of learning to learn is clouded by a plethora of phrases used to describe the kind of thinking under discussion, so much so that Cuban (1984) described the area as 'a conceptual swamp' (p 676). Thus there is need for clarification of terms and to decide the similarities and differences between, for example: higher order thinking skills; problem-solving; critical thinking; reasoning; decision-making; metacognition; productive thinking; creative thinking; mindful thinking. Further, the meaning of these terms has changed over time and according to whether the context is philosophical or psychological (Lewis and Smith, 1993). For example, Lewis and Smith's (1993) discussion of the definition of higher order thinking notes that critical thinking has in the past been regarded as synonymous with problem-solving; it was later regarded as evaluation or judgement, and at times it is treated as a combination of evaluation and problem-solving. Lewis and Smith argue that the term 'higher order thinking' encompasses problem-solving, critical thinking, creative thinking and decision-making. They define it as follows: 'higher order thinking occurs when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations' (Lewis and Smith, 1993: p 136).

However, Newman (1990) points out that what may require higher order thinking by some may only require lower order thinking by others and so 'to determine the extent to which an individual is involved in higher order thinking, one would presumably need to know something about the person's intellectual history' (p 45).

Fogarty and McTighe (1993) have identified three phases in attempts to define and teach thinking skills. These are: skill acquisition (e.g. Marzano and Hutchins, 1985), critical and creative thinking as required for problem-solving (e.g. de Bono, 1983) and 'thoughtful application' (Brown and Palinscar, 1982). It seems clear that for decision-making and lifelong learning, as required in a rapidly changing world, it is the third kind – thoughtful application – that is required.

Adey (2001) describes the structure of thinking skills as being layered. One layer is a general processing ability. This is variously described as intelligence, working memory, formal operations, multivariate thinking, higher order thinking skills. In a different layer, there are particular thinking skills of which critical and creative thinking are seen by Adey as 'key nodes'. He identifies critical thinking as analytical thinking, a convergent type of thinking, related to reasoning and to formal and informal logic. Creative thinking he describes as divergent, lateral – the type of thinking that generates new ideas. Other skills, which can be regarded as derivatives of these skills are evaluative thinking, exploratory thinking and hypothesising.

Adey, like Fogerty and McTighe, sees problem-solving as a combination of all other thinking skills, calling on some more than others according to the demands of the problem. This means that problem-solving is not readily defined, since it will in some circumstances draw more on convergent critical thinking and, in others, on divergent creative thinking. Thus, where problem-solving is the focus of assessment, the thinking skills involved will depend on the particular problem.

Ennis (1993, 1996) also defines a critical thinker in terms of convergent skills, such as the ability to judge the credibility of sources, the quality of argument, to identify conclusions, reasons and assumptions, to develop and defend a position on an issue, to plan experiments and judge experimental design. Huitt (1998) agrees with this in offering the definition of *critical thinking* as 'the disciplined mental activity of evaluating arguments or propositions and making judgments that can guide the development of beliefs and taking action' (p 2). He also gives a useful definition of *non-critical thinking*, as thinking which does not consider current data but relies on past practices, or brainstorming that involves saying whatever comes to mind without evaluation. *Non-critical thinking* also includes *creative thinking*, which involves synthesis and requires 'an individual to look at parts and relationships (analysis) and then to put these together in a new and original way' (Huitt, 1998).

Whilst this discussion has added little to the clarity of the conceptual field, it does lead to the identification of critical and creative thinking as key factors, or nodes (in Adey's terms), in thinking. There appears to be a convergence of views that these are the two pillars of higher order thinking; indeed Lewis and Smith's definition (above) is close to a combination of critical and creative thinking. Lawson, Abraham and Renner (1989, p 27) list thinking skills under the following main headings which, again, echo earlier definitions of critical thinking:

1. Skill in accurately describing phenomena
2. Skill in sensing and stating causal questions
3. Skill in recognising, relating and stating alternative hypotheses and theories
4. Skill in generating logical predictions
5. Skill in planning and conducting controlled experiments to test hypotheses and theories
6. Skill in collecting, organizing, and analysing relevant experimental and correlational data

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

9

7. Skill in drawing and applying reasonable conclusions

All of these are described as 'skills', defined as 'the ability to do something well' (Lawson, 1993, p 171). Others have dropped the word 'skill' in favour of referring to 'thinking'. In the context of the current review, where the concern is with assessment, it is necessary to seek evidence of thinking in what students do, say and write or in how they perform. Thus it is not inappropriate, although not essential, to use the term 'skill'. What is more important is to be clear that we are concerned with outcomes of the kind listed by Fogarty (1992), as these are closely aligned with what is required by students for dealing with choices, problems and challenged in everyday life and for continuing to learn from experience and from their own thinking. We are, therefore, using the phrase 'creative and critical thinking skills' to convey the aspects of thinking whose assessment is the subject of this review.

In summary the definitions of creative and critical thinking used in this review are as follows:

- **critical thinking:** involving the evaluation of arguments or propositions in relation to evidence, reasoning, drawing conclusions
- **creative thinking:** involving relating together principles, ideas, information and entities in new and original ways to generate new entities or ideas

1.2.2 ICT and assessment

Some of the potential advantages of using ICT for assessment over traditional methods are as follows:

- There exists the use of a greater variety of images (including moving images) than is possible in paper and pencil testing.
- Students may interact with the material in flexible and creative ways.
- There is the possibility of interaction of students with other people online, synchronously or asynchronously, thus making communication 'for real'.
- Where appropriate, students can be provided with large amounts of data from which they select what is relevant for a particular problems or decision.
- Real-life situations are modelled in which problems are posed requiring manipulation of variables to find optimum conditions.
- The assessment of creativity is undertaken through specialized programs that facilitate effective presentation.
- The pace of assessment can be controlled either by the students or by the program, depending on whether immediacy of response is a significant aspect.
- It is technically possible to develop programs to mark open responses and short essays reliably.
- Self-assessment can be readily provided.
- Peer-assessment can be organized online, again, either synchronously or asynchronously.
- Students can obtain feedback from their teacher or other students
- Assessment can more easily be provided on the basis of 'test when ready' for individuals rather than at a fixed time for all.

A key feature of ICT that supports these advantages for assessment is the interactivity that is afforded. Interactivity means that there are options provided by, and feedback from, the technology; that is, the technology is used for more

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

10

than a convenient or novel replacement for a paper-based technique. Although ICT includes the use of film, video, audio-tape, CD Rom and DVD as well as computer-based technology, currently the technology that provides for interaction is mainly computer-based. Thus, in this review, the search was made for all types of technology but excluded those that do not allow for interaction. One of the intentions of this review is to provide information about the extent to which any of these potential advantages are being realised in practice and used in extending assessment to creative and critical thinking skills.

1.3 Policy and practice background

Considerable value is currently placed on the development of the abilities to think and to learn as goals of education. The reasons most commonly advanced for this relate to the need to prepare young people for life and work in a rapidly changing society in which they will have to make more choices than did those living in past decades (Claxton, 1999). The value of teaching thinking skills is supported by research evidence recently reviewed by McGuinness (1999). The ability to continue learning throughout life is acknowledged as essential for future generations and thus has to be a feature in the education of every student. Developing this ability involves learning how to learn, the achievement of various cognitive, affective, and conative outcomes needed to provide the skill, the will, the flexibility in thinking, and the energy needed to make effective decisions. This is recognised as needed by students in all developed countries, as underlined by the OECD:

> 'Students cannot learn in school everything they will need to know in adult life. What they must acquire is the prerequisites for successful learning in future life. These prerequisites are of both a cognitive and a motivational nature. Students must become able to organise and regulate their own learning, to learn independently and in groups, and to overcome difficulties in the learning process. This requires them to be aware of their own thinking processes and learning strategies and methods.' (OECD, 1999, p 9)

The predominant methods in regular use for assessment today are paper and pencil based. As a consequence, according to McFarlane (2001), 'the curriculum is assessed almost exclusively in terms of that which can be tested through paper and pencil tests' (p 232). Where computers are used, they tend to mimic paper and pencil tests (McFarlane and de Rijke, 1999). They have for some time been used for assessing recall and simple applications (Bunderson *et al.*, 1988). There is, however, a growing number of examples of ICT use that go beyond these boundaries and show that the use of ICT can become important for assessing creative and critical thinking. For example, Klieme (2000) describes how computer simulation can be used for assessing complex problem-solving. He points out the following:

> '...these instruments provide a new quality of problems, distinguished by high levels of complexity and, in particular, by a dynamic character. These dynamic tasks have three advantages over static paper-and-pencil tasks:

> (a) The demands of the tasks are enhanced by an active search and continuous processing of external information and feedback. In solving written problem-solving tasks, it is also possible to apply, evaluate and – if necessary – modify processing strategies. Interaction with the computer, however, makes such a course of action inevitable.

(b)   In this medium, the problem situation can be made much more authentic than in a written test.

(c)   Not only results, but also the course of the problem-solving process can be recorded and assessed, i.e. the type, frequency, length and sequence of interventions made by the subjects. This provides process-based indicators of problem-solving strategies.'
(Klieme, 2000, p 8)

The use of computers in this way is, however, not without its problems; the disadvantages as well as the advantages need to be explored. Linn (1994) has pointed out some of the problems of performance assessment, which apply equally to these uses of computer-based assessment. Russell and Haney (2000) have found evidence that scores on computer-based assessments are influenced by the level of keyboard skills. When compared with handwritten answers, the computer-based results could be higher or lower depending on keyboarding speed.

In theory, using ICT in assessment should have much the same advantages as its use in teaching and learning. In some cases, the distinction between the two is blurred. McFarlane (2001) cites an example of students creating a multimedia presentation in the context of a drugs education programme. 'The added value offered by the technology was to assist learners to externalize a representation of their own thinking and understanding of a set of complex ideas, which subsequently revealed their misconceptions to themselves and colleagues in ways which traditional methods had not' (p 232).

1.4 Research background

The systematic review of research on summative assessment and testing by Harlen and Deakin Crick (2002) was focused on the impact on students' motivation for learning. However, one of the sub-questions addressed by the review was to consider the evidence of impact on teachers and teaching in those studies where this was reported in addition to the impact on students. The relevant findings were as follows:

- When passing tests is high stakes, teachers adopt a teaching style which emphasises transmission teaching of knowledge, thereby favouring those students who prefer to learn in this way and disadvantaging and lowering the self-esteem of those who prefer more active and creative learning experiences.
- External tests have a constricting effect on the curriculum, resulting in emphasis on subjects tested at the expense of creativity and personal and social development.
- High stakes tests often result in a great deal of time being spent on practice tests and the valuing of test performance and undervaluing of other student achievements.
- Teachers' own assessment becomes summative in function rather than formative
- Teachers can be very effective in training students to pass tests even when the students do not have the understanding or higher order thinking skills that the tests are intended to measure.
- Teachers can influence students' self-assessment criteria towards learning processes.
(Harlen and Deakin Crick, 2002, p 56)

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

12

These findings are entirely consistent with those of other reviews in this area, such as those by Crooks (1988), Linn *et al.* (1982) and Shepard (1991). Kellaghan *et al.* (1996) have expressed doubts about the aims of the education reform which emphasises that higher level thinking and problem-solving skills are compatible with the programmes of high-stakes testing. The research they reviewed also undermined the claim that better tests will lead to better teaching and learning within a high-stakes testing ethos. This was supported by the evidence reported by Gordon and Reese (1997), that teachers can train students to pass any kind of test, even those intending to assess higher thinking skills, frustrating those who consider that teaching to well designed tests can influence teaching in positive directions (e.g. Yeh, 2001).

Thus there is extensive evidence that teachers are influenced by what is tested and that what is not tested tends to be given minimal attention. However there is also a warning here that, as long as the stakes are high, extending the range of tests may not increase teaching towards the intended learning outcomes, but rather may increase teaching more directly to what is tested.

There is, however, equally strong evidence that, when the purpose of assessment is formative (that is, to help learning), then performance is very substantially improved. Black and Wiliam's (1998) review of classroom assessment found assessment to have a wholly positive impact on learning when it includes certain key features which emerged from the studies. These features include careful attention to questioning, particular forms of feedback, the involvement of students in self- and peer-assessment and the use of assessment in modifying teaching. Black and Wiliam acknowledged that such practices require large shifts in teachers' perceptions of their roles in relation to their students, but that considerable gains in achievements are possible as a result. Recently, Black *et al.* (2002) have published examples of how these practices can be implemented. They also note ways in which summative tests can be used formatively. However, these ways apply 'where teachers have control over the setting and the marking, but their application may be more limited for tests where the teacher has little or no control' (Black *et al.*, 2002, p 13).

This research into assessment suggests that, if creative and critical thinking outcomes are to be achieved, then these outcomes need to be included in assessment of all kinds and for both formative and summative purposes. Inclusion in summative assessment and testing, on an equal footing with other learning outcomes, is likely to ensure their place in the curriculum. Formative assessment will ensure that appropriate learning experiences are provided; indeed, many of the outcomes described as creative and critical would appear to be encouraged by the process of formative assessment itself.

1.5 Authors, funders, and other users of the review

The reason for proposing and conducted this review, as indicated above, was the importance that is attached to thinking skills in the information age. The authors, both based at the Graduate School of Education of the University of Bristol, completed in 2002 a review of research into the impact of testing and summative assessment on the motivation for learning of students aged 4 to 18 (Harlen and Deakin Crick, 2002). This revealed, as noted earlier, that teaching followed the tests, which were narrowly focused, paper-based, and largely neglected creative and critical thinking. At the same time, the development of computer-based technologies provided the potential for assessment of the thinking skills widely

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

13

deemed essential for today's learners. It was therefore felt important, at this time, to identify evidence on the impact of using ICT for assessing critical and creative thinking skills, and to make this evidence available to those involved in developing policy and practice in assessment.

It was anticipated that the findings would inform policy-makers concerned with the development of tests and examinations, researchers studying the use of ICT in teaching and learning, and advisers concerned with teacher professional development. Outcomes of the review, relevant to these potential users, were communicated to them in appropriate forms.

The review was funded solely by the contract between the DfES-funded EPPI-Centre at the Institute of Education and the University of Bristol on behalf of the Assessment and Learning Research Synthesis Group (ALRSG). Members of the ALRSG and overseas advisers are listed on page i. The review was carried out by the authors with the guidance of the review group and the participation of members in various stages, as noted in sections 2.1, 3.3 and 4.5.

1.6 Review questions

The overall question addressed in the review was:

- What evidence is there of the impact of the use of ICT for assessment of creative and critical thinking skills on students and teachers?

In order to achieve all the aims of the review, it was necessary to address the further subsidiary questions:

- How does any impact vary with the experience of the students and the conditions of the assessment?
- How does any impact vary according to whether the purpose of the assessment is formative or summative?
- What are the implications for assessment policy and practice of these findings?

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

14

2. METHODS USED IN THE REVIEW

This chapter describes the systematic methodology of the review and the role of potential users in it. It presents the inclusion and exclusion criteria used in searching for studies and identifies the subsequent stages of the process of this EPPI-Centre review.

2.1 User involvement

The ALRSG includes the following users: a primary school headteacher, who became an LEA primary link inspector during the course of the review; a secondary school deputy headteacher with responsibility for assessment; and an LEA Adviser (Head of School Improvement and Lifelong Learning) who is a member of the Association of Inspectors and Advisers for Assessment. Through the members of ARG, who form the core of the ALRSG, there are links with the DfES Assessment for Learning Group, the Association of Inspectors and Advisers for Assessment, and the National College for School Leadership. The international expert advisers provide links with institutions overseas.

The members of the ALRSG were closely involved at all stages of the review: in defining the review question; drawing up the protocol; discussing inclusion and exclusion criteria; keywording (four members, in addition to the authors, took an active part in this); and in data-extraction, following a training session in using the EPPI-Reviewer. The group discussed the map of included studies (see section 2.3.1) and helped to develop the framework for synthesising findings (see section 2.3.4).

At the stage of draft findings, a half-day conference was held with a wider range of users involved in research, policy and teaching related to the use of ICT in education, including assessment. The conference, held at the Graduate School of Education (GSOE) of the University of Bristol was attended by 20 staff and higher degree students. These included members of staff leading research projects in ICT and members of their research teams, members of the Centre for Assessment Studies and several students and tutors on the MSc course in Technology, Education and Society, which includes members of the Institute for Learning and Research Technology. The conference discussed a summary of the findings of the review and the implications of the review for policy, practice and research (see section 5.5).

2.2 Identifying and describing studies

2.2.1 Defining relevant studies

The following criteria were established to guide the decisions about which studies to include.

(i) Language of report
Studies included were written in English. Although it was possible to translate from other modern European languages, the search strategy dealt with

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

15

databases and journals in the English language and studies in other languages were not actively sought.

(ii) Focus of assessment and testing

Studies were included if they reported formative or summative assessment and testing of creative and critical thinking (as defined in the background to this review), using ICTs.

(iii) Type of study and study design

For the purposes of addressing the overall review question, studies were only included if they evaluated the processes involved in assessment and testing of critical and creative thinking skills using ICT (process evaluations) or the impact of assessment and testing using ICT on students, or teachers (outcome evaluations). Outcome evaluations employing a range of study designs were included: studies that compared the impact of different testing programmes or ICT-based testing with non-ICT-based situations (trials) or those that studied one group over time before and after the introduction of ICT-based testing (one group pre- and post-study). Process evaluations, which employed a range of study designs and methods of data collection, were included (e.g. cross-sectional surveys, in-depth interviews, observation).

For the purpose of setting out the background and context of the review, a wider range of studies was drawn upon, such as existing reviews, descriptions of practice and discussion of relevant issues in using ICT for assessment of creative and critical thinking.

(iv) Setting and population

Studies were only included if they reported on pupils in school or pre-school, between the ages of 4 and 18. Studies dealing only with students outside this age range, in further or higher education or adult education, were excluded.

(v) Date of research

Initially no date limit was set on the studies sought. However, since the search of online databases showed that most studies involving the use of ICT for more than routine knowledge testing were published since 1990, the handsearching of journals was limited to this date.

2.2.2 Search strategy

A list of likely journal and database sources was created from references in key studies already obtained and building on previous reviews. The search for studies meeting the inclusion criteria was begun by searching bibliographic databases and registers of educational research. Details of the search strategies for electronic database are given in Appendix 2.1. Relevant journals that are online were also searched by computer. Other journals held in the library were searched by hand, as were back numbers of those only recently put online. Details of journals hand-searched are given in Appendix 2.2. Study titles and abstracts were reviewed before being entered into the database. Other studies were found by scanning the references lists of already identified reports, making requests to members of relevant associations and other review groups, and using personal contacts.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

16

2.2.3 Applying inclusion and exclusion criteria

The titles and abstracts of all studies were screened against the inclusion and exclusion criteria in section 2.2.1, before being entered into the database of identified studies. Thus screening of titles and abstracts was a one-stage process, followed later by further screening of full texts.

2.2.4 Characterising included studies by keywording

Full texts of studies were obtained where possible and screened using the core set of 10 keywords of the EPPI-Centre Educational Keywording Sheet and a set of four review-specific keywords (see Appendix 2.3). Discussion in the Review Group resulted in refinement of an initial set of review-specific keywords and a revised set was used in the final mapping. Keywording was carried out by the authors and four members of the Review Group.

Keywording resulted in the exclusion of a number of papers, for reasons that were documented and are indicated in Figure 3.1. The agreed keywords for the remaining studies were used to produce the systematic map of included studies.

2.2.5 Quality assurance processes

Records were made of all searches: electronic database searches were trialled, repeated and documented; dates of journals searched were recorded. Where necessary, judgements about inclusion and exclusion criteria were checked by both researchers. Twenty studies were keyworded by two people to highlight and eliminate differences in interpretation of definitions. EPPI-Centre staff also carried out a quality assurance role in keywording a sample of studies.

2.3 In-depth review

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

It was decided that all studies in the descriptive map would be included in the in-depth review. Following keywording and mapping, the studies included were closely read for data-extraction and evaluation of the weight of evidence in relation to the review question.

2.3.2 Methods for extracting data from studies in the in-depth review

The selected studies were entered in the EPPI-Reviewer 2.0. This enabled information to be extracted using the *Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research* (EPPI-Centre, 2002) and review-specific questions. The EPPI-Reviewer was used to code the results of the EPPI-Centre keywording strategy and review-specific keywords. The data-extraction guidelines required answers to be given to around 100 questions and details of specific aspects of studies to be quoted or summarised. This included judgements about the methodological quality of the studies to be assessed. The review-specific questions required judgements to be made about the weight of

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

17

evidence of each study in relation to the review question, as described in more detail in the next section.

2.3.3 Assessing quality of studies and weight of evidence in relation to the review question

In order to ensure that conclusions were based on the most sound and relevant evidence, judgements were made using the EPPI weight of evidence criteria. This involved judgements about three aspects of each study and the combination of these to give an overall judgement of the weight that could be attached to the evidence from a particular study to answer the review question.

The three aspects and their combination are outlined in A to D below.

A: Soundness of methodology
Judgement of how well the study had been carried out was informed by the responses to the following eight aspects of the study in the section of the EPPI Reviewer relating to quality of the study:

- Is the context of the study adequately described?
- Are the aims of the study clear?
- Is there an adequate description of the sample used in the study and how the sample was identified and recruited?
- Is there an adequate description of the methods used in the study to collect data?
- Is there an adequate description of the methods used in the study to analyse data?
- Is the study replicable from this report?
- Do the authors avoid selective reporting bias?

In turn, the answers to these questions were informed by data about, for example, the methods for ensuring the validity of data analysis identified in an earlier part of the data-extraction process. The judgement of methodological soundness was thus dependent on what was reported in the study. The lack of information about a certain feature did not necessarily mean that this feature was not attended to in practice by the study, only that it was not reported by the author of the study. Studies were rated as high (H), medium (M), or low (L) in relation to methodological soundness, according to the extent to which these criteria were met.

B: Appropriateness of research design 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 review question. In theory, some study types or designs might be better matched to the focus of the review than others. This was not a judgement on the value of the study in its own right, but only in respect of how well its design enabled the review question to be answered. Studies were rated high, medium and low in relation to this aspect.

C: Relevance of the particular focus of the study for answering the review questions
As in B, this judgement concerns the match of the study to the purposes of the review and is not a judgement on the value of the study *per se*. In this case, the aspect of interest is the topic focus of the study: that is, how well the nature of the

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

18

data collected helped to answer the reviews questions. Again the judgements were in terms of high, medium or low relevance.

D: Overall weight that can be given to the evidence in relation to the review focus

The judgements for the three aspects were combined into an overall weight of evidence towards answering the review question. In doing this, where there was a difference of judgement between A, B and C, the overall judgement was based on the majority rating or on an 'average' (e.g. ratings of all of L, M and H would be M).

2.3.4 Synthesis of evidence

In the review question, the nature of the use of ICT in assessment was unspecified, although the actual medium of ICT used was categorised in the keywording process. Similarly the nature of the impact on students and teachers was unspecified. Both of these have emerged from the analysis of the selected studies.

Consideration of the meaning of critical and creative thinking indicates that, in order to assess these skills, students have to be given the opportunity, for instance,

- to access and evaluate information for a particular purpose
- to engage in authentic-rational thinking
- to make links between ideas
- to evaluate conclusions and support their reasoning with arguments
- to create new understandings, ideas or entities

Moreover, in order to know to what extent the intended thinking processes have been taking place, it is necessary for the assessment to provide information about students' processing and not just about their learning outcomes.

In relation to the nature of the use of ICT for the assessment of creative and critical thinking skills, as it is reported in these studies, there were different 'affordances' of ICT which were exploited by the researchers. The term 'affordances' is used within a theoretical framework that sees ICT as a set of cultural tools offering opportunities for action (Barnes, 2000; McFarlane, 2002; Wertsch, 1998).

The affordances of ICT in the assessment of creative and critical thinking skills can be categorised into at least two broad genres. These are the assessment of the *processes* of creative and critical thinking and the assessment of *outcomes* which are presumed to have come about as a result of the processes of creative and critical thinking skills. These *outcomes* can be further broken down into (a) 'self-report' measures that assess individuals' dispositions, values and attitudes on the basis of items designed for the purpose of assessing creativity and critical thinking, and (b) learning outcomes which themselves could not have been achieved without the processes of creativity and critical thinking on the part of the learner.

The processes of creative and critical thinking skills have to do with the learner's capacity, respectively, to think laterally; to utilise intuition, imagination and generate new ideas; to think analytically; to make meaning across data and within experience; to make causal connections between data; and to create new

knowledge (see section 1.2). These processes involve affective, conative and cognitive aspects of learning and often include social and collaborative learning. Information and communication technologies (ICT) are able to capture, report on and facilitate these processes, when used for both summative and formative assessment purposes. They also capable of enabling other new forms of assessment. It is worth noting that the use of a computer program for assessment involves a teacher (assessor), either directly or indirectly, as well as a learner. The computer is a cultural tool that can be used in the service of learning and development. Whether the teacher receives feedback directly from the computer or mediated through the learner, it is still feedback that has been humanly constructed and thus represents the values and intentions of the teacher/assessor, located in a particular socio-historical context.

Some of the key affordances of ICT for the assessment of the *processes* and *outcomes* of creative and critical thinking *indicated in these studies* are identified here: firstly the speed and convenience of using ICT for assessment and feedback, rather than the time and effort required by similar measures using pencil and paper; and, secondly the capacity of ICT to capture and report on the representation of complex dynamic processes which would otherwise be very difficult to monitor and assess.

The interaction of students and teachers with assessment processes using ICT can be categorised as information-collection, information-tracking, information-processing and information-feedback. Information-collection via a computer can include visual representations of learners' thinking and processing of information (as in mind maps), including the content of learners' concepts and the links the learner makes between concepts. It can include the collection of social learning interactions, synchronous and asynchronous, between learners as through networked learning scenarios, using email and other forms of communication. It can include collection of data about learners' problem-solving decisions in a virtual or semi-virtual learning environment. The non-linear, visual, dynamic and multi-media nature of ICT affords the collection of learning data which are relevant to the deployment of creative and critical thinking skills.

Information-tracking via a computer means that complex processes, which would otherwise be opaque to learners and teachers, can be recorded and tracked over time. Thus the changes and direction of a learner's creative and critical thinking can be mapped, and this information can be used in various ways.

Information-processing via a computer of creative and critical thinking skills includes the capacity of the computer to analyse and calculate across a range of inputs, to compare inputs – for example, novice versus expert mind maps – and to evaluate against a rubric.

Information-feedback can be to either learner or teacher/assessor. It can be in the form of prompts, expert versus novice representations, or the direction of processes and thinking pathways. It can be summative in intent or formative.

In Figure 2.1, a model is proposed in an attempt to simplify this analysis for the purposes of this review. The impact of the use of ICT *as observed in these studies* is divided into the impact on teachers and the impact on learners. For learners, the impact is on their attainment in critical and/or creative thinking – often both, as required in problem-solving. Outcomes reported that are inferred rather than specifically based on evidence are offered as possible explanatory factors under 'implications'. These include outcomes such as motivation, strategic

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

20

awareness of learning processes, meaning-making and self-regulation. For teachers, the impact relates to the nature and the type of information that is available to them in relation to the processes and outcomes of creative and critical thinking. For teachers also there are implications for, rather than substantive evidence of, a recognition of the need for more professional knowledge about learning, especially creative and critical thinking skills.

2.3.5 In-depth review: quality assurance procedures

At the stage of data-extraction, each of the selected studies was considered independently by at least two people, one being one of the researchers. The members of the Review Group had been trained in using the *Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research* (EPPI-Centre, 2002) in the EPPI-Reviewer 2.0. Comparisons were made between pairs of responses entered into the EPPI-Reviewer by each person involved in extracting data. Reconciliation of differences was face to face between those involved in five instances and in the other seven, the two researchers examined the complete print-outs of the responses and arrived at agreed responses through checking against the original study. EPPI-Centre staff also carried out a quality assurance role in data-extraction for a sample of studies.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

21

Figure 2.1: Affordances, processes and outcomes of using ICT for the assessment of creative and critical thinking skills emerging from the studies

Some affordances
- Tracking and representation of learner processes over time and cumulatively
- Simulation of real-life scenarios for decision-making and problem-solving
- Visual, dynamic, non-linear data collection, analysis and feedback
- Assessment of learning outcomes, which are presumed to have required creative and critical thinking skills
- Speed of processing and feedback
- Memory extension
- Development of self-awareness of learning and motivation for learning
- Development of competence and motivation for learning

Engagement of students and teachers with assessment processes using ICT
- Information collection – space, time, connections, multiple forms of evidence
- Information tracking – time and direction
- Information processing – analysis, computation, comparison and evaluation
- Information feedback – prompts, expert versus novice representations, evaluation of processes and outcomes

Impact on LEARNER
- Achievement / Performance

Impact on TEACHER
- Better knowledge of students' processing and learning
- Summative knowledge of student attainment

Implications
- Teachers' recognition of need for more professional knowledge about learning, especially of creative and critical thinking skills
- Possible impact on learners': motivation, awareness of learning processes, meaning-making and self-regulation as explanatory factors

3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS

This chapter presents the results of the procedures described in Chapter 2 for identifying studies. It provides a diagrammatic overview of the filtering of studies that took place in selecting the studies that gave sound and relevant evidence in relation to the review question. Maps of initially selected studies discussed by the Review Group and the final selection are given.

3.1 Filtering papers from searching to map to synthesis

The result of the search process was a set of 103 studies which were identified as relevant and entered into an EndNote database (Db1). This was a one-stage process, since titles and abstracts were screened in relation to inclusion and exclusion criteria before being entered into the database. Because a large amount of the searching was completed by hand, perusing complete journal articles, the initial list of citations is smaller than may be expected. Of the 103 studies, the full texts of 94 were obtained. The full texts were screened using the criteria as set out in Table 3.1.

Sixty-two studies were excluded by the application of these criteria, and labelled accordingly, before keywording. Thirty-two studies were categorized using the EPPI-Centre keywording sheet and the review-specific keywords (see Appendix 2.3). This led to the exclusion of a further 20 studies. Thus 82 studies in total were excluded. In some cases, two or more criteria applied; hence the total for the various criteria exceeds the number excluded.

Table 3.1: Label codes for exclusion criteria at full text screening stage

Code	Criterion
A	Not formative or summative assessment
B	Not ICT-based
C	Not school, aged 4 to 18
D	Not creative or critical thinking skills
E	Not research
BR	Background and recommendations

Of the 82 excluded studies, 13 were considered to be of sufficient relevance to be placed in a separate database for use in background discussion and possible guidance in relation to recommendations. The remaining 12 (i.e. 94-82) provided the basis for the map, and a consultation with the Review Group about the inclusion of studies for the review.

Figure 3.1 summarises the number of studies at each stage of the review.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

23

Figure 3.1: Filtering of papers from searching to map to synthesis

1. Identification of potential studies

One-stage screening: papers identified in ways that allow immediate screening

2. Application of inclusion/ exclusion criteria

Potential included N =103

Papers not obtained N =9

Full document screened N =94

Papers excluded N =82

Criterion* A: N = 24

Criterion* B: N = 6

Criterion* C: N = 17

Criterion* D: N = 11

Criterion* E: N = 42

3. Characterisation

Systematic map studies included N =12

4. In-depth review

In-depth review N =12

*Note that these criteria are not mutually exclusive

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

24

3.2 Characteristics of the included studies based on core and review-specific keywords

Characteristics of the 12 included studies are summarized in Table 3.2. Further analysis of the characteristics of the studies is given in the tables in Appendix 3.1.

As might be expected, with only 12 studies, the occurrence, or absence, of patterns is difficult to discern. Two thirds of the studies had a curriculum focus on science and no other subject area was well represented. This is contrary to the usual emphasis in assessment, which frequently favours mathematics and English. It may suggest a lack of attention to the potential of ICT for assessing creative and critical thinking in these commonly assessed subjects. The fact that 11 of the 12 studies were evaluations of researcher-manipulated situations perhaps indicates that ICT is not yet commonly used in assessment of creative and critical thinking skills, or possibly that these skills are not widely included in the outcomes of education that are regularly assessed.

The studies covered a wide range of creative and critical thinking skills, with problem-solving predominating. With this range, it is not possible to report on an impact of assessing particular skills.

Whilst the majority of studies concerned assessment that provided information on student achievement to the teacher, half also provided formative information to the student. By definition, this feedback is intended to help learning, thus the same student tasks were both assessment and learning tasks.

Table 3.2: Characteristics of the 12 studies selected for data-extraction

General EPPI-Centre keywords	Number (N=12)
Identification of reports	
Citation	1
Handsearch	5
Electronic database	6
Status of reports	
Published	12
Country where study carried out	
USA	5
England	4
Israel	1
Germany (US researchers and students)	2
Topic focus**	
Assessment	12
Curriculum	9
Equal opportunities	1

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

25

Table 3.2: Characteristics of the 12 studies selected for data-extraction (continued)

Curriculum focus***	
General	1
History	1
Maths	1
Science	6
Population focus	
Learners	12
Age of learners	
5-10	4
11-17	8
Sex of learners	
Females and males	12
Educational setting of the study	
Secondary (including middle school in US)	8
Primary	3
Nursery	1
Type of study	
Evaluation: naturally occurring	1
Evaluation: researcher-manipulated	11

Review-specific keywords	**Number (N=12)**
Type of ICT used**	
Multi-media	1
Computer program	12
Intranet/Network	4
Creative and critical thinking skills assessed**	
Problem-solving	5
Communication skills	3
Access and assessing information	3
Metacognitive reflection	2
Application of knowledge	4
Affective skills	2
Generation of new ideas or products	1

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

26

Table 3.2: Characteristics of the 12 studies selected for data-extraction (continued)

Review-specific keywords	Number (N=12)
Assessment information provided**	
Formative feedback to teacher	6
Formative feedback to student	9
Summative for teacher	9
Summative for student	0
Peer-assessment assessment	0
Impact reported**	
On students' motivation	3
On students' achievement	8
On teachers' teaching	2
On teachers' perception of students	4

** These categories were not mutually exclusive.
*** Only relevant to the nine studies where the topic focus included the curriculum.

3.3 Quality assurance results

For the 20 studies key-worded by two people there were four instances of disagreement, all concerning the type of study. These were resolved through discussion amongst review group members. EPPI-Centre staff carried out a quality assurance role in keywording a sample of studies.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

27

4. IN-DEPTH REVIEW: RESULTS

This chapter describes the nature and findings of the finally selected studies. The synthesis of findings in relation to the main review question is given in three sections dealing with the three types of impact identified in the studies. Findings addressing the subsidiary question are also given and other impacts that were reported by authors as explanatory factors are summarised.

Due to the focus of the review covering both summative and formative assessment, the findings from the selected studies cover aspects of both the assessment and the learning aspects of ICT. The findings concerning the learning aspects of ICT should not be considered as an exhaustive summary of this area, as systematic searches were not completed for learning in ICT.

4.1 Description of included studies

An outline of the aims, study type, findings and conclusion of 12 selected studies is given in Appendix 4.1.

Table 4.1 gives the weight of evidence each study is judged to provide, based on the criteria for soundness of method, appropriateness of the research design and relevance of the topic for answering the review question. Only two of the studies were judged to give evidence of high weight. This is an indication of the need for more and better research that is concerned with the assessment of creative and critical thinking skills.

Table 4.1: Results of assessment of weight of evidence for each study (H = High; M = Medium; L = Low)

Study	A Soundness of method	B Appropriateness of research design	C Relevance of topic focus	D Overall weight of evidence
Cifuentes L , Yi-Chuan J (2000) *Concept learning through image processing*	M	M	L	M
Jackson B (1989) *A comparison between computer-based and traditional assessment tests, and their effects on pupil learning and scoring*	M	M	L	M
Kumar D *et al.* (1993) *Effect of HyperCard and traditional performance assessment methods on expert-novice chemistry problem-solving*	M	H	M	M
Littleton K *et al.* (1998) *Gender, task scenarios and children's computer-based problem-solving*	H	M	L	M
Masterman L , Sharples M (2002) *A theory-informed framework for designing software to support reasoning about causation in history*	M	M	H	M
McFarlane A *et al.* (2000) *Assessment and multimedia authoring - a tool for externalising understanding*	L	M	H	M
Means B *et al.* (2000) *Developing assessments for tomorrow's classrooms*	L	L	H	L

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

28

Study	A Soundness of method	B Appropriateness of research design	C Relevance of topic focus	D Overall weight of evidence
O'Neil H *et al.* (1999) *Computer-based collaborative knowledge-mapping to measure team processes and team outcomes*	H	M	M	M
Osmundson E *et al.* (1999) *Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings*	H	H	H	H
Schacter J *et al.* (1997) *Feasibility of a web-based assessment of problem-solving*	M	M	M	M
Schacter J *et al.* (1999) *Computer-based performance assessments: a solution to the narrow measurement and reporting of problem-solving*	M	M	M	M
Tzuriel D, Shamir A (2002) *The effects of mediation in computer assisted dynamic assessment*	M	H	H	H

4.2 Synthesis across studies: overall review question

What evidence is there of the impact of the use of ICT for assessment of creative and critical thinking skills on students and teachers?

In the main question addressed in the review, the nature of the impact on students and teachers was unspecified and was allowed to emerge from the analysis of the selected studies. Table 4.2 shows the impacts and outcomes that were reported. The types of impact fell into three categories as identified in Figure 2.1. Analysis of the outcomes reported by each study identified three main types of outcome. Table 4.2 lists the occurrence of these three outcomes for each of the 12 studies. In several cases, study authors gathered other information, more or less informally, and reported these as tentative findings or as speculations about the cause of a particular impact on students. This was particularly the case in relation to affective outcomes, mentioned where students were reported as liking working on a computer or where the impact of feedback to the students could be affective as well as cognitive. These are reported in section 4.4 as possible explanatory factors.

The findings relating to the main review question will be discussed under three categories as shown in Table 4.2:

- teachers' summative knowledge of students' achievement
- teachers' knowledge of students processing and learning
- reported impact on students' achievements

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

29

Table 4.2: Type of study and impact/outcomes reported

Study	Study type	Reported outcome for teachers		Reported impact on students' achievement
		Summative knowledge of students' achievement	**Knowledge of students' processing and learning**	
Cifuentes L , Yi-Chuan J (2000) *Concept learning through image processing*	Evaluation: research-manipulated		X	X
Jackson B (1989) *A comparison between computer-based and traditional assessment tests, and their effects on pupil learning and scoring*	Evaluation: researcher-manipulated	X		X
Kumar D *et al*. (1993) *Effect of HyperCard and traditional performance assessment methods on expert-novice chemistry problem-solving*	Evaluation: researcher-manipulated	X		X
Littleton K *et al*. (1998) *Gender, task scenarios and children's computer-based problem-solving*	Evaluation: researcher-manipulated	X		
Masterman L and Sharples M (2002) *A theory-informed framework for designing software to support reasoning about causation in history*	Evaluation: researcher-manipulated and description		X	
McFarlane A *et al*. (2000) *Assessment and multimedia authoring - a tool for externalising understanding*	Evaluation: naturally-occurring	X		X
Means B *et al*. (2000) *Developing assessments for tomorrow's classrooms*	Evaluation: researcher-manipulated	X		X
O'Neil H *et al*. (1999) *Computer-based collaborative knowledge-mapping to measure team processes and team outcomes*	Evaluation: researcher-manipulated	X		
Osmundson E *et al*. (1999) *Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings*	Evaluation: researcher-manipulated	X	X	X
Schacter J *et al*. (1997) *Feasibility of a web-based assessment of problem-solving*	Evaluation: researcher-manipulated	X		X
Schacter J *et al*. (1999) *Computer-based performance assessments: a solution to the narrow measurement and reporting of problem-solving*	Evaluation: researcher-manipulated	X		X
Tzuriel D, Shamir A (2002) *The effects of mediation in computer assisted dynamic assessment*	Evaluation: researcher-manipulated		X	X

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

30

4.2.1 Studies where the outcome is a summative assessment of some aspect of creative and critical thinking

Nine of the 12 studies had as an outcome the summative assessment of some aspect of creative and critical thinking, as shown in Table 4.3. One of the studies, by Osmundson *et al.* (1999), provided evidence of high weight relating to this review; seven studies provided evidence of medium weight and one of low weight.

Table 4.3: Studies where the outcome is a summative assessment of some aspect of creative and critical thinking

Study	Study focus	Outcomes reported	Age group	Country	Overall weight of evidence
Jackson B (1989)	Effect of computer-based assessment on motivation and learning	Score on conceptual understanding in science test, computer-based compared with paper-based	10 – 14 Middle school	England	M
Kumar D *et al.* (1993)	Effect of computer-based assessment compared for 'expert' and 'novice' students	Score on test of balancing chemical equations, computer-based compared with paper-based	11 – 17 High school	USA	M
Littleton K *et al.* (1998)	Gender differences in solving computer-based problems	Performance of girls and boys in different problem-solving tasks	11 – 12	England	M
McFarlane A *et al.* (2000)	Use of multimedia package to assess students' understanding of drug use	Students' understanding judged from their multimedia presentation and from written test	5 – 10	England	M
Means *et al.* (2000)	Trial of program for assessing problem-solving	Students' use of the web, reasoning with information and communication	5 – 16	USA	L
O'Neil H *et al.* (1999)	Assessment of teamwork skills through collaborative knowledge-mapping	Team processes and problem-solving as measured by knowledge-mapping	11 – 16 Middle and high school	Germany (by US researchers)	M
Osmundson E *et al.* (1999)	Use of computer-based knowledge-mapping for assessing and developing understanding	Pre- and post-test of problem-solving as measured by knowledge-mapping scores of control and experimental groups	5 – 10	USA	H

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

31

Study	Study focus	Outcomes reported	Age group	Country	Overall weight of evidence
	of the human body				
Schacter J *et al.* (1997)	Use of concept-mapping to assess processes of problem-solving in environmental studies	Scores on processes involved in problem-solving and concept maps scores	11 – 16 Middle and high school	USA	M
Schacter J *et al.* (1999)	Use of concept-mapping to assess impact of using web on problem-solving in environmental studies	Pre- and post-test scores on concept-mapping as a measure of problem-solving in environmental studies	11 – 16 Middle school	Germany (with US researchers)	M

The study by Osmundson *et al.* was one of four studies, selected in this review, which used computer-based concept/knowledge-mapping as a measure of conceptual understanding or problem-solving. The focus on conceptual understanding or problem-solving depended on the nature of the mapping task, but both types involved critical and creative thinking. These studies all derive from work at the Centre for Research on Evaluation, Standards and Student Testing (CRESST) at the University of California, Los Angeles, where concept-mapping has been developed as a means of assessing aspects of understanding that are not assessed by conventional assessment methods. Initially the concept-mapping was on paper, but Osmundson *et al.* observed (1999, p 3), 'The incorporation of technology into the knowledge-mapping tasks has provided new instructional and assessment opportunities, by providing the capacity for real-time and database-driven storage, retrieval, scoring, feedback, reporting, Internet/Web access, and increased interactivity. These capabilities enable new and novel uses of knowledge maps for instruction (e.g. as a tool to focus student discussion over the Internet) as well as for assessment (e.g. immediate scoring and reporting, analyses of students' map over time)'.

Concept, or knowledge, mapping involves creating links between words representing concepts in order to produce propositions about their relationship. The link is made by an arrow, which is labelled with the nature of the link. A list of concepts may be given or created by the mapper; similarly the link terms may be given or created. In the Osmundson *et al.* (1999) study, where maps were concerned with the interconnections within and between systems of the human body, both concepts and link terms were given, having been chosen by a group of teachers, experts and researchers so as to give reasonable propositions appropriate to the age of the students. Examples of concepts are 'brain' and 'body', which can be linked by 'controls' to give 'brain controls body', the direction being indicated by an arrow. The concepts 'brain' and 'body' can also be linked to other words, creating a complex map of interconnected concepts.

In the computer-based form, students could select concepts and links terms by selecting from a list and clicking and dragging across the screen. For a completed map, the computer lists all the propositions created. During development of the mapping tool, these propositions were scored, by expert raters, for quality, and

the ratings were stored in a database. Each completed map is given various scores relating to the interconnections among systems. These scores, produced automatically using algorithms, took into account other features, such as the number of relevant propositions and their quality. Thus the propositions in completed maps could be automatically scored for various features.

The objective of the study by Osmundson *et al.* (1999) was 'to investigate how computer-based knowledge-mapping could be used simultaneously as an instructional tool and an assessment tool in a classroom setting' (p 1). Using a non-equivalent control group design, the researchers collected data for two groups over a period of six weeks. Students in both groups created individual knowledge maps in week 1 and week 6 as pre- and post-tests. As an additional post-test, all students were also asked to write an essay to assess their understanding of the respiratory, digestive and circulatory systems of the human body. In the intervening weeks, the experimental students worked in groups and created three knowledge maps, receiving feedback from the computer on their performance, whilst the control group worked in small groups used a computer and other materials to research the human body. Thus the knowledge-mapping was used both for assessment and as a learning experience.

The findings indicated no significant differences between the control and experimental groups on the pre-test map performance. All data suggested the groups were comparable. All students improved on all measures (number of pragmatic propositions, number of scientific propositions and number of highly principled propositions and interconnection) and overall the experimental group showed a higher mean performance on the outcome variables than the control group. The inter-correlations between knowledge map content score and holistic essay score were reported as being of 'moderate magnitude', reaching significance at the 0.05 level. For each knowledge map measure, an analysis of variance was conducted. Significant differences were found between control and experimental groups on the content score, interconnection score, number of scientific propositions score and number of highly principled propositions, all in favour of the experimental group. There was no significant difference for the number of nonsense propositions.

Microanalysis of one group's propositions was used to investigate how students' propositions changed over time. The total number of propositions created increased from one occasion to the next supporting the conclusion that 'the process of learning appeared to involve adding bits and pieces of knowledge to their system of understanding' (p 21). Videotapes of the group working indicated that the visual display of the monitor and its interactive nature 'provided a tangible visual organiser around which students could negotiate meaning' (p 21).

The authors conclude that the change in mapping scores from pre-test to post-test for all students, both experimental and control, showed that the knowledge assessment measures were sensitive to the intervening instruction and were therefore effective in providing information about achievement. 'Tools such as the knowledge-mapper, with their automated scoring and ability to deliver immediate feedback to users, have the potential to serve both as instructional supports for students and teachers and as information providers for teachers about the ways in which students gradually come to refine their understandings in science' (p 23).

In this study, the computer was used to provide feedback of the kind that could not be provided by other means, since scoring individual concept or knowledge maps is too time-consuming to be a realistic approach to providing feedback

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

33

during learning. This would also apply to using knowledge-mapping as a summative measure. The study provides evidence that automatic scoring can be applied to computer-based knowledge-mapping, thus making it feasible as a summative assessment of students' critical thinking about complex relationships.

Three other studies is this group (O'Neil *et al.*, 1999; Schacter *et al.*, 1997; Schacter *et al.*, 1999) were concerned with the use of concept / knowledge-mapping as an outcome measure in the context of investigating the impact of various conditions on understanding or problem-solving.

O'Neil *et al.* (1999) used knowledge-mapping to measure the outcome of team processes during a group knowledge-mapping task. In a pre-post design, they arranged students to work in team of three, separated from each other and communicating using networked computers. To collaborate the students in a team sent messages to each other; most messages were predefined and could be sent by pressing buttons. The messages were classified as relating to five aspects of collaboration: adaptability, co-ordination, decision-making; interpersonal and leadership.

Performance was measured by how well a team's map compared with experts' maps. The results of the investigation did not show the expected relationship between measures of teamwork processes and the performance of a team as measured by the knowledge map produced, either for the pre-test or the post-test. 'Where significant relationships existed, the direction of the relationship was opposite of what we expected. In general, the relationship between teamwork processes and teamwork outcome measures was negative suggesting that the more messages a team sent the poorer they performed on the knowledge-mapping task' (p 19).

The authors observed that the task was highly knowledge-dependent and participants lacked the requisite knowledge to engage with each other at a substantive level. Further, the use of messages may have deflected attention from the substantive discussion about content on to the procedural aspects of constructing a knowledge map and selecting messages.

Schacter *et al.* (1997) used concept maps as an outcome measure in order to investigate processes related to concept-mapping. This was also a pre-test / post-test design but, in this case, after the pre-test concept-mapping, the students worked individually using a simulated worldwide web to improve their concept maps. Their performance whilst doing this was used to explore how the processes of exploration, extraction, simplification and organisation contribute to problem-solving performance. Processes used in information-seeking were recorded by logging electronically every mouse click and source visited on the web. The log was used to give a measure of each student's performance in exploratory information-seeking. The number of relevant web documents accessed (bookmarks) was recorded, as was the degree to which students focused on areas of the map where they needed more knowledge. The improved maps constituted the post-test.

It is implied that the students completed these activities in a single session and so it can be assumed that the significant change in concept-mapping score that was found was due to the use of the simulated web environment. A regression analysis with final concept-mapping scores as the independent variable and information-seeking variables as the dependent variable revealed that none of these variables predicted scores in final concept-mapping. However, there was a

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

34

relationship found between extracting more relevant information (relevant bookmarks) and final concept map performance. Students' scores on a test of metacognition were not found to predict higher final scores on concept-mapping.

A similar lack of relationship was found for scores on 'focusing' on concepts where more information was needed or accessing feedback on students' performance. The main finding from this study concerned methodology; that is, 'the feasibility of automated collection, scoring, and reporting of students complex problem-solving process data and using that data to make inferences about problem-solving performance. As we have shown, process data can be collected in inexpensive and non-time intensive ways. Further, detailed reporting of students' cognitive processes while engaged in learning activities is a feasible goal when using networked technologies' (Schacter *et al.*, 1997, p 19).

The study by Schacter *et al.* (1999) used a similar design and some of the same measures as an earlier study by Schacter *et al.* (1997). Students were pre- and post-tested using concept-mapping relating to environmental science, but in this study, pre-testing took place in the fall of one year and post-testing in the following spring. After the pre-test, students searched a simulated web for information to improve their maps and, whilst doing so, could have access to feedback about their performance. In this study, emphasis was laid on the feedback: 'Outcome feedback presented students with information describing whether concept-link-concept propositions in their concept maps were correct and where additional work was needed. Feedback was based on comparisons of student's and expert's performances in creating knowledge maps (p 410).

The pre- and post-tests scores on concept maps were compared; the groups were matched but not identical. The findings were that students not only improved their concept-mapping significantly but also their problem-solving processes used in searching, browsing, accessing feedback, etc. Data collected over the year showed that this improvement was not due to direct teaching in environmental science and so the authors concluded that 'student problem-solving processes played an importance role in increased problem-solving performance' (p 414). In contrast with Schacter *et al.* (1997), in this study, process measures of browsing, focused browsing and accessing feedback were significantly correlated with concept-mapping performance. Searching almost reached statistical significance but bookmarking did not predict mapping performance. However, it was noted that bookmarking was not extensively used.

The authors emphasised that their result demonstrated the feasibility of collecting and reporting to students on their performance and processes in real time using computer technology. They commented that 'Not only did we report to students how they performed on outcome measures, but also which thinking processes contributed to or detracted from their performance' (p 415). They concluded that the feedback improved students' performance.

Jackson (1988) explored the question of whether there was an advantage for students in taking a computer-based test compared with a test on paper. He used a computer to administer multiple-choice science tests to middle school students. During the tests, the students received feedback on their performance. Their performance was compared with that of a control group who took the same tests on paper. The students were randomly assigned to the two groups. Tests were given regularly over a period of about two months at approximately fortnightly intervals.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

35

The outcomes were assessed by a final paper-based test including all the previous questions and a cloze procedures test. The computer group scored significantly higher than the control in the multiple-choice test and all students had low scores in the cloze test. The author concluded that the computer feedback produced higher final scores than those not receiving feedback. Unfortunately this study did not distinguish between the effect of feedback and the effect of the computer. Thus the main finding of this study was that using the computer both to test and to give feedback raised the level of performance of the students.

Although the effects on motivation and attitudes were included in the intentions of the project, there was no formal data-collection relating to these outcomes. The claim that motivation was increased by use of the computer was based on the students' 'enthusiasm for doing a test during breaks and lunch times' (p 814).

Kumar *et al.* (1993) also compared performance in computer-based (HyperCard) and paper-based tests, in this case with older students involved in balancing chemical equations. The computer provided immediate feedback to the students and a record for the teacher. The design was to compare two forms of the assessment both for students who were 'expert' in chemistry (used to solving problems for themselves and preparing for lab experiments by themselves) and those who were novices (had less experience of chemistry in depth than the 'experts'). The novice and expert groups were randomly assigned to using either the HyperCard or paper-based method of balancing chemical equations. All students had been taught about chemical equations four weeks before the trial.

The students were scored on correctness and a 'performance score', which combined correctness with the difficulty level of the equation and the number of attempts made to balance it. The findings showed that performance and correctness scores for both 'expert' and 'novice' students were significantly higher on the HyperCard method than on the paper-based method. The performance scores of novices using the HyperCard method nearly equalled the performance score of the experts using the paper-based method. 'Time on task' for the experts decreased from paper-based to the HyperCard method while 'time on task' for the novices increased from the paper-based method to the HyperCard method.

The authors considered that the higher performance with the HyperCard method may have resulted from:

- the non-linear environment of the HyperCard method;
- the ease of using the mouse interface as compared with using a pen for solving equations;
- the immediate feedback provided by the computer program;
- the computer itself providing an added 'external memory for the student while balancing the equations thereby reducing the cognitive demand on working memory' (p 7).

The authors also commented that the feedback motivated the students to stay on-task, although no other evidence of motivation was collected to support this conjecture.

MacFarlane *et al.* (2000) compared students' understanding shown by a conventional test with what emerged from analysis of multimedia presentations created by the students. The students were from schools involved in a multimedia project. They developed a multimedia presentation on matters related to drug

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

36

use. After producing their multimedia presentation, they were given a test and the data about their understanding of drugs from the two sources were compared. The students were also interviewed for qualitative findings.

The results indicated that whilst one third of the students answered all the test questions correctly, and a third responded correctly to all but one question, their multimedia presentations and their discussion about them showed little understanding of the key concepts. The comparisons between the test and the analysis of the students' multimedia presentation and their talk about it 'suggests that the test results incorrectly gave an impression that the pupils had a thorough grasp of key facts. As a method of assessing pupil achievement, the test failed to reveal a number of assumptions and associations, and preconceptions and misunderstandings, that were evident in the pupils' multimedia work and talk about that work' (p 204). The authors conclude that multimedia authoring was effective in assessing students' real understanding of the complex issues involved.

The study by Littleton *et al.* (1998) has relevance for this review in relation to the factors that may influence student performance when using computers to solve problems. The study, involving 11- and 12-year-old students, used two versions of a computer-based problem-solving program. In 'King and Crown' the quest takes the form of the retrieval of a crown from an island. The 'Honeybear' task is designed to offer exactly the same formal task demands but within a different scenario. Both tasks present the same problem in the same adventure game format and call for an identical solution strategy. After being given instruction in using the program, the students worked individually. There were two studies, each involving equal numbers of boys and girls, and comparing performance in the same way. In the second study, a different version of King and Crown was designed, called 'Pirates'. In each case half the boys and half the girls were randomly assigned to one or other of the scenarios. Children were also given an attitude test immediately after the problem-solving.

The results for both studies were that problem-solving performance was significantly affected by the version of the task that was used, the influence being particularly strong for the girls. In the second study 'the performance of the boys remained relatively unaffected by the software type, whereas the performance of the girls is far superior when using the **'Honeybears'** version' (p 336). The results of this study 'offer clear evidence that it is the difference in the scenario, as opposed to any confounding differences in the characteristics of the interface, which are accounting for these effects' (p 337). The attitude test added little information because all students expressed strongly positive attitudes. However, there was a significantly stronger liking for computers in general expressed by boys compared with girls.

Such gender differences have, of course, been found in conventional paper-based tests, although not necessarily testing problem-solving. However, it is possible that the imagery used in the computer-based tasks has a stronger impact and exaggerates the gender difference. Indeed the authors suggest that the girls identified with the characters in the software more than the boys and suggest that this factor may have influenced the girls' motivation to engage with the task. They also suggest that the level of confidence girls have in using computers may be more important than gender per se.

The study by Means (2000) is rated as providing low weight evidence since it describes the trial and evaluation of a prototype assessment task for problem-

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

37

solving and is essentially work in progress. The amount of detail given about the trial and the study type and the design fall short of providing dependable evidence, but the focus of the study is highly relevant to the review. The task is an extended one involving students searching a range of websites for relevant information and assessing the credibility of what they find. The task poses the problem that a group of foreign exchange students wants to come to the US for the summer and needs to choose one of two cities. In the middle school version and the high school version of the task, different criteria are given as a basis for decision. Students were asked to evaluate the credibility of the information on particular web pages and to formulate a search query for finding additional, relevant information. Students were asked to compare and weigh all this information in making their choice and present reasons for their choice in writing to the foreign students. The whole takes about two hours to complete. The student outcomes assessed included technology use, reasoning with information, and communication. Their responses were scored by two independent raters.

The findings of the trial indicate that the students had no difficulty in using the search and word-processing facilities. They demonstrated greater proficiency at finding topically appropriate information than at reasoning with the information or communicating conclusions in a well-organised and thoughtful manner. There were relationships between students' prior experience with technologies and their scores. The study shows that high-level skills involved in accessing, assessing and communicating information can be assessed by computer, and that this assessment can be embedded in learning activities. The authors discuss the value of this approach for formative assessment. 'Technology can make assessments of the kinds of skills needed for the 21st century knowledge economy more feasible – providing assessment tasks that mimic the real-world problems' (p 5).

Summary of main findings relating to summative assessment

The relevant finding from the one study providing evidence of high weight was as follows:

- Computer-based concept-mapping with automated scoring can be used to provide summative assessment of creative and critical thinking about complex relationships.

The findings from studies giving medium weight-evidence, each from one study only, were as follows:

- Use of computers to assess teamwork did not provide evidence that aspects of collaboration results in increased problem-solving as measured by computer-based knowledge-mapping (O'Neil *et al.*, 1999).
- Automated collection and scoring of the processes used in problem-solving provide additional information relevant to problem-solving performance (Schacter *et al.*, 1997).
- Using a computer both to test and give feedback to students can increase the level of performance, compared with students taking the same tests on paper (Jackson, 1989).
- Using a computer-based program for solving problems of balancing chemical equations enables students to achieve a higher level of performance than solving equations on paper (Kumar, *et al.*, 1993).

- A conventional multiple-choice test can give a false impression of understanding compared with the analysis of multimedia presentations on the same topic (McFarlane *et al.*, 2000).
- Subject matter used in computer assessment of problem-solving affects the outcome for girls more than boys (Littleton *et al.*, 1998).

4.2.2 Studies where the outcome is increased knowledge for the teacher of students' processing and learning

Table 4.4 gives information about the four of the 12 selected studies where a main purpose of the assessment was formative; that is, it provided information for the teacher about students' learning that could potentially be used to adapt teaching and improve students' learning. Two of the studies provide high-quality evidence, the other two providing medium-quality evidence.

Table 4.4: Studies where the outcome for the teacher is greater knowledge of students' processing and learning

Study	Study focus	Outcomes reported	Age group	Country	Overall weight of evidence
Cifuentes L and Yi-Chuan J (2000)	The effect of the computer-based manipulation of images	Conceptual learning in science of students receiving different interventions	11 – 16 Grades 7 and 8	USA	M
Masterman L and Sharples M (2002)	Assessment of causal reasoning in history	Feasibility of using computer program to assess students' achievement in causal reasoning	11 – 14	USA	M
Osmundson E *et al.* (1999)	Use of computer-based knowledge-mapping for assessing and developing understanding of the human body	Pre- and post-test of problem-solving as measured by knowledge-mapping scores of control and experimental groups	5 – 10	USA	H
Tzuriel D and Shamir A (2002)	To examine the effects of computer assisted dynamic assessment compared with assessment by examiner only	Scores on test of seriational thinking of groups experiencing combinations of pre-testing mediation with or without computer	5 – 10 Kinder-garten, aged 6	Israel	H

The study by Osmundson *et al.* (1999) has already been described in section 4.2.1. Two of the research questions of the study were: 'Did knowledge-mapping support and/or promote learning?' and 'How was the knowledge-mapper used during instruction?' Evidence suggesting a positive answer to the first of these questions came from the greater change in performance scores for the experimental group which used the knowledge-mapper in the intervention phase compared with the control group which used other resources to improve learning. The second question was addressed by using information provided by the computer program to follow the changes in the propositions students created in

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

39

their maps. 'Our analysis suggests that students developed their ideas in four fundamental ways: (a) new ideas (nonsense, pragmatic, scientific, and highly principled – meaning that greater understanding of interconnections was shown) were added to students' repertoire of understandings; (b) some nonsense or impossible ideas were transformed into more scientific understandings, or they disappeared from the maps; (c) some pragmatic ideas became more principled; and (d) more connections between systems were developed' (p 22). The computer-mapping tool provided feedback both to the students and the teacher about the success in developing propositions during the intervention phase. The authors claimed that 'by using the mapping tool in the classroom, a teacher can facilitate children's learning and also gain a better understanding of that learning by seeing how students are grasping new material and where instructional emphasis could/should be placed' (p 23).

The study by Tzuriel and Shamir (2002) compared the effect of using a computer program that provided feedback to student and teacher during learning with the impact of feedback from the teacher only. Kindergarten children aged 4 to 7 were assigned to two groups, both being assessed on some seriation tasks before and after some teaching which was either assisted by a multimedia computer-based program or was given by a teacher/mediator only. Some children used the computer for the pre- and post-tests, whilst for others these were administered by the instructor- the objects to be placed in order being displayed on cards rather than the screen. The design combined the different test forms with the mediation forms in various ways to assist interpretation of the cause of any differences. The focus of the study was the difference between the computer-plus-mediator and mediator-only intervention between pre- and post-tests. This intervention was described as 'dynamic assessment', which was essentially formative assessment since information was used to help the children process information to solve the seriation problems.

The computer-assisted dynamic assessment was a program called 'Think-in-Order' based on test of *Children's Seriational Thinking Modifiability* that was also used as pre- and post-test. The program guides the mediator to apply certain principles ('scaffolding' and 'graduated prompt') in the teaching process. The use of the program depends on the mediator; it cannot be used by children on their own. 'The assessment procedure is managed mainly by the program, leaving the mediator to deal with aspects that are beyond the programmed mediation. The mediator can respond flexibly to the evolving needs of the learner during the whole assessment process' (p 24). In the mediator-only dynamic assessment, the mediator was the only source of feedback to the student.

The seriation tasks were problems that required arrangement of objects on a continuum whilst controlling for one or more dimensions that are embedded within the same set of stimuli: for instance, pictures of cars (different numbers, sizes and darkness of colour). The cars were given in mixed order at the bottom of the screen and the children asked to order them at the top by dragging and dropping. Icons represented the three possible task dimensions and the children had to click the one they used. If the children made a mistake, they received a graduated prompt, increasing in explicitness until full definition of the first dimension was provided.

The study results showed that all groups made gains from pre- to post-test but the gains were greater for the children who experienced computer-assisted mediation. For the group for which both tests and mediation involved the computer, the change in score was highly significantly greater than for the group

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

40

not using the computer at all. The pre-test to post-test gain for the group tested and mediated without the computer was similar to the gain for the whole group experiencing teacher mediation (for some of whom the tests were administered by computer), suggesting that the test mode did not account for the difference. Neither was any difference in teaching time found to account for the difference between groups. Other analyses showed that the more difficult the task, the greater the superiority of the computer-assisted mediation. The authors eliminated alternative interpretations and were able to conclude that the computer-assisted dynamic assessment had a greater impact on performance than teacher-only dynamic assessment.

Cifuentes and Yi-Chaun (2000) studied the effects of training seventh and eighth grade students to generate and use computer graphics. They aimed to study how the manipulation of images during knowledge construction in science can support learners' meaning-making and conceptual development, and how the students' graphical representations can inform teacher feedback to students. A post-test only design was used, with students being randomly assigned to experimental and control groups. The experimental group participated in a workshop of three 50-minute training sessions on manipulating and generating computer graphics as a help to studying. The workshop aimed to help them to recognise and represent the underlying structure of text, relate new concepts to prior knowledge, highlight distinctive features and use graphics for review. The group was then given some print-base material to study (on a science concept relating to the general properties of matter). The control group was also given the same material to study. After 50 minutes for studying this material, students kept both their study notes and reading material. The next day, they were tested on the science material, handing in their notes prior to taking the test.

The authors found no difference in favour of the experimental group in any of the measures used. They suggested that this might be because of the immaturity of the middle-school students involved, who needed more guidance in using effective visualisation and were 'extremely distracted' by the computer software. However, they claimed that 'We discovered that learner generated graphic representations of concepts provided a rich resource for students' teachers. Representations of learners' understanding provided teachers with a way of knowing whether or not students grasp concepts. Teachers suggested that if students cannot visualize the concept, perhaps they don't thoroughly understand the concept' (p 81).

The study by Masterman and Sharples (2002) was a formative evaluation of computer program designed to support reasoning about causation in history. It also used graphic representation. One of its aims was to find if a computer program could be used to identify aspects of students' thinking useful to teachers. The authors developed a computer program designed:

> 1. To stimulate pupils to refine their understanding of causation and its associated concepts through constructing and interacting with diagrammatic representations of the causal relationships of specific historical situations.
> 2. To make visible the capabilities and limitations of pupils' causal reasoning through a) completed representation and b) records of the actions taken by pupils in constructing their representations' (p 175).

The development of the software was informed by a socio-cognitive and socio-constructivist theoretical framework. This was developed through the use of the

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

41

concept of the 'zone of proximal development'. The concepts of modelling-coaching-fading, and the 'conversational metaphor' were also utilised.

In the program, entitled 'Storm Ahead' the pupils created diagrams linking 'cause boxes' (different causal factors) by arrows to the final event. The diagram constituted, in graphical form, the student's explanation of the causes of the event. When using the program, students start by creating a cause box and assigning properties to it (long or short term, religious, economic, etc). When they have created more than one box, they draw arrows to denote relationships. The computer records a log of their moves.

In the formative evaluation of the program, it was used by four classes of 12- and 13-year-olds 'as the culmination of their study of the cause of the English Civil War' (p 179). After instruction in using the program, the students spent about an hour on their diagrams, working either individually or in pairs. In addition to the log files, diagrams and pupil questionnaires, data were collected about the processes of using the computer program and about pupils' and teachers' views on usability, easiness, etc.

The program was easy or very easy to use by most of the pupils (over 92%). Almost 90 percent considered it easier to draw a diagram with the program than using paper and pencil and most were willing to use the program again in another history topic. The Log file showed few procedural errors but they proved less useful than expected in showing how the computer supported the students' learning.

From the analysis of the students' diagrams, it appeared that most seemed able to distinguish between causes and mere facts. 84 percent of the diagrams contained links and, on average, 69 percent of the boxes were linked to at least one other box. However, fewer than one-third of links could be classified as causal in nature and the prevalence of fragmented chains indicated that pupils had great difficulty linking causes into any coherent structure. Most pupils were able to distinguish between causes and mere facts. However, pupils had difficulty linking causes into any coherent structure; that is, they judged each cause in isolation from the other causes.

It was concluded that analysis of the diagrams developed in Storm Ahead revealed shortcomings in students' causal reasoning and indicated the program's potential for 'diagnosing fallacies in pupils' cause reasoning, such as their inability to distinguish between causal and non-causal relationships' (p 183). In considering a more active pedagogical role for the computer, however, they warned that 'where the program provides coaching it must avoid misleading pupils into thinking that their reasoning is necessarily 'wrong' (p 183).

Summary of main findings relating to formative assessment

The two studies providing high-weight evidence showed that the computer can provide the teacher with help by:
- storing and recording information about how students are developing understanding of new material (Osmundson *et al.*, 1999);
- taking over some of the role of assessing and providing feedback to students so that the teacher can focus on ways of supporting learning that are beyond the program's mediation (Tzuriel and Shamir, 2002).

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

42

Findings from the studies giving medium-weight evidence, each from one study only, indicate that:

- the performance of middle school pupils was not improved by training in the use of computer graphics aimed at helping them make meaning of new material (Cifuentes and Yi-Chuan, 2000);
- a program designed to support students' causal reasoning through diagrammatic representation provided the teacher with useful information about their thinking through analysis of the diagrams but not from the log files of their computer moves (Masterman and Sharples, 2002).

4.2.3 Studies where the impact is on the students' achievement

Eight of the 12 studies reported an impact of using ICT for assessment on students' achievement in critical thinking or creative thinking. Two of these were the studies providing evidence of high weight, with the other six providing evidence of medium weight, as indicated in Table 4.5.

Table 4.5: Studies where the outcome for the pupil was their performance

Study	Study focus	Outcomes reported	Age group	Country	Overall weight of evidence
Cifuentes L and Yi-Chuan J (2000)	Effect of the computer-based manipulation of images	Conceptual learning in science of students receiving different interventions	11 – 16 Grades 7 and 8	USA	M
Jackson B (1989)	Effect of computer-based assessment on motivation and learning	Score on conceptual understanding in science test, computer-based compared with paper-based	10 – 14 Middle school	England	M
Kumar D *et al.* (1993)	Effect of computer-based assessment compared for 'expert' and 'novice' students	Score on test of balancing chemical equations, computer-based compared with paper-based	11 – 17 High school	USA	M
McFarlane A *et al.* (2000)	Use of multimedia package to assess students' understanding of drug use	Students' understanding judged from their multimedia presentation and from written test	5 - 10	England	M
Osmundson E *et al.* (1999)	Use of computer-based knowledge-mapping for assessing and developing understanding of the human body	Pre- and post-test of problem-solving as measured by knowledge-mapping scores of control and experimental groups	5 – 10	USA	H

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

43

Table 4.5: Studies where the outcome for the pupil was their performance (continued)

Study	Study focus	Outcomes reported	Age group	Country	Overall weight of evidence
Schacter J (1997)	Use of concept-mapping to assess processes of problem-solving in environmental studies	Scores on processes involved in problem-solving and concept maps scores	11 – 16 Middle and high school	USA	M
Schacter J *et al.* (1999)	Use of concept-mapping to assess impact of using web on problem-solving in environmental studies	Pre- and post-test scores on concept-mapping as a measure of problem-solving in environmental studies	11 – 16 Middle school	Germany (with US researchers)	M
Tzuriel D and Shamir A (2002)	To examine the effects of computer assisted dynamic assessment compared to assessment by examiner only	Scores on test of seriational thinking of groups experiencing combinations of pre-testing mediation with or without computer	5 – 10 Kinder-garten, aged 6	Israel	H

All these studies have been described above, so we refer here specifically to the findings relating to students' performance assessed in the studies. The two studies providing high-weight evidence both used computer-based test material as a measure of change and as the means of intervention between pre- and post-test. In the study by Osmundson *et al.* (1999), the performance of students provided with feedback during computer-based knowledge-mapping was compared with students who used other means to improve their understanding of the systems of the human body. The gain in scores was significantly higher for the experimental group, using the knowledge-mapping, than for the control group. Tzuriel *et al.* (2002) used the computer to provide feedback to young children who were at the same time receiving help from the teacher with some seriation tasks of the kind used in the tests. These children made greater gains in test scores than a control group who were helped to solve similar problems, between pre- and post-test, only by the teacher.

Two studies (Jackson, 1989; Kumar *et al.*, 1993), providing medium-weight evidence, reported increased performance as a result of feedback from the computer during computer-based problem-solving or knowledge-mapping.

Three other studies in this group were concerned with the impact of using computer technologies for an intervention between pre- and post-test that was designed to improve performance but was not the same as the test material. Cifuentes and Yi-Chuan (2000) reported no impact of training 11- and 12-year-olds to use computer-graphic packages to help with their study. Schacter *et al.* (1997) reported no impact of using a web environment, with feedback on performance to search for relevant information on the immediate performance in concept-mapping. However, Schacter *et al.* (1999) found significant effects on

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

44

post-test concept-mapping with a longer time between pre- and post-test and with feedback on performance.

The study by McFarlane *et al.* (2000) showed that assessing understanding through the use of computer technology to produce presentations revealed a poorer grasp of basic concepts than appeared to be the case from performance on a multiple-choice test of the same material.

Summary of main points relating to impact on performance

- Findings from two studies providing high-weight evidence supported by two giving medium-weight evidence indicate that feedback from the computer during the use of test material improves performance in later use of the same test material (Jackson, 1989; Kumar *et al.*, 1993; Osmundson *et al.*, 1999; Tzuriel and Shamir, 2002;).

- There is conflicting evidence from two studies relating to the impact on performance of using the web to search for information (Schacter, 1997; Schacter, 1999).

- One study showed that training in using computer a graphics packages to assist study did not improve performance (Cifuentes and Yi-Chuan, 2000).

- In one study, the use of multimedia authoring to create presentations revealed limitations in true understanding of complex issues that were not revealed by conventional multiple-choice tests (McFarlane *et al.*, 2000).

4.3 Synthesis across studies: the subsidiary review questions

- How does any impact vary with the experience of the students and the conditions of the assessment?
- How does any impact vary according to whether the purpose of the assessment is formative or summative?
- What are the implications for assessment policy and practice of these findings?

The first two of these questions will be considered in this section. The third question is addressed in section 5.5.

4.3.1 Impact of factors relating to the conditions of testing and the experience of students

The conditions investigated in some of the selected studies were teamworking, expertise in the subject matter and familiarity with, and liking for, the use of computers.

O'Neil *et al.* (1999) studied the impact of interaction among team members working at a distance on a joint knowledge map on their collaboratively produced map. The medium-weight evidence from this study did not support the expectation that certain aspects of teamwork would be associated with better

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

45

performance in the knowledge-mapping test. One possible reason suggested by the authors was the distraction caused by the demands of sending messages. Since Osmundson *et al.* (1999), in a study giving evidence of high weight, found that groups worked well together in improving their maps and this had a positive impact on their later scores in knowledge-mapping, it appears that the element of teamworking at a distance was not enabling real collaboration in the task. Another suggestion related to the poor background knowledge of the students in the environmental science that was assessed.

Subject-matter expertise was studied by Kumar *et al.* (1993) in their comparison of computer-based and paper-based methods of balancing chemical equations. Equal numbers of students regarded as 'expert' chemistry students and regular, or 'novice', chemistry students were randomly assigned to one or other testing method. The medium-weight evidence from this study was that both expert and novice students achieved higher scores in the computer-based method, the advantage being particularly large for the novice students, whose computer-based scores were nearly equal to scores of the experts using the paper-based method.

The impact of the test medium was also investigated by McFarlane *et al.* (2000) in a study providing medium-weight evidence. They found that requiring students to use multimedia technology to create a presentation of their ideas revealed deficiencies in their understanding that were not shown in a conventional test.

Means *et al.* (2000) reported a relationship between performance in an extended problem-solving task and students' prior experience of the computer technologies they had to use in solving it, including internet searching. This evidence was of low weight in relation to this review. Means *et al.* (2000) did not report on any gender difference in this experience, but there was medium-weight evidence from the study by Littleton *et al.* (1998) that girls expressed less liking for computers than boys. This could lead both to less experience and less confidence in using computers. Littleton *et al.* (1998) noted that studies with adults showed that confidence in using the technology might be more important than gender as such, and that this might apply to younger children also.

Main points
- Working with a team at a distance has been found not to improve performance in knowledge-mapping (O'Neil *et al.*, 1999; medium-weight evidence).
- Computer-based assessment can enable students to achieve at a higher level than in equivalent paper-based tasks (Jackson, 1989; medium-weight evidence).
- Using technology can probe students' understanding to a greater degree than conventional tests (McFarlane *et al.*, 2000; medium-weight evidence).
- Students' experience with computers and attitude towards them can influence computer-based test performance (Means *et al.*, 2000; Littleton *et al.*, 1998; medium-and low-eight evidence).

4.3.2 Impact of the and aims of the assessment programme

The feature of assessment using ICT most frequently commented upon as having an impact on students and teachers was the affordance of feedback, a feature of assessment with a formative purpose.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

46

The most common form of feedback to students arises from their interaction with the computer and the visual changes they make in the display. In some cases, the feedback goes no further than this immediate effect of changes or links made and shown on the screen. The study of Masterman and Sharples (2002), providing weight-evidence, is an example of the computer taking this neutral role. The authors considered, however, that making the links 'visually salient' to the students helped their learning. This view was supported by Kumar *et al.* (1993), also providing medium-weight evidence, who suggested that the immediate feedback showing the outcome of students' attempts to solve chemical equations encouraged them to stay on-task until a satisfactory solution was reached.

In other studies providing medium-weight evidence, the computer had a more active role in giving feedback, by providing comments, scores, or information about processes and performance from which the students could infer what they needed to do to improve. Feedback of a generally encouraging kind was provided in the study by Jackson (1988). Students who did not give a correct answer on their first attempt were given a second chance and, if incorrect again, would then be given the correct answer. They were, therefore, able to know correct answers for each question as they proceeded. The students taking tests under these conditions performed significantly better on a final test compared with the control group taking the tests on paper with no feedback.

Feedback based on the processes used in a task requiring higher level thinking was provided in the studies of Osmundson *et al.* (1999) and Schacter *et al.* (1999) through computers logging the moves made. Osmundson *et al.* (1999) provided high-weight evidence that the use of a knowledge-mapping test program as an instructional tool, providing immediate feedback to students about how their maps 'scored' in comparison with an 'expert' map, increased these students' later knowledge-mapping test scores, compared with students who used other resources for improving their knowledge of the subject. Schacter *et al.* (1999), providing medium-weight evidence, gave feedback to students on their thinking processes as well as outcomes during concept-mapping, and concluded that this was effective in improving students' performance over time. In a study of similar design, but over a shorter timescale, also giving evidence of medium weight, Schacter *et al.* (1997) used a knowledge-mapping program to provide feedback to students and teachers about the thinking processes as well as problem-solving performance and suggested that computer programs make feasible the use of process data for assessment purposes.

A key feature of programs is the choice of content or context for the assessment task. Any assessment of a skill has to sample performance using a small number of contexts and specific topics taken from a potentially large pool of possibilities. When the assessment takes the form of an extensive task, such as may be required for valid assessment of problem-solving, any bias associated with the choice of subject or context will have an impact on performance. Only one medium-weight evidence study, by Littleton *et al.* (1998) reported on the impact of the topic, focusing on possible gender bias that might be associated with it. The study compared performance of boys and girls on problem-solving tasks presented through different scenarios, expected to be of different interest to boys and girls. There was a significant difference in girls' performance according to the scenario, whilst there was little impact of this variable on boys' performance. The study's design eliminated other interpretations and the authors pointed to the importance of the findings for software designers.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

47

Main points
- Six studies (one providing high-weight evidence and five medium-weight evidence) show that interacting with a computer provides feedback that supports better performance, even if this only reflects back to the students the moves and links made by students in a visual representation of relationships.
- Computers can also provide information about processes in reaching a solution that gives additional feedback to students and teachers (Schacter *et al.*, 1997; Schacter *et al.*, 1999; medium-weight evidence).
- There is evidence that the choice of content influences the performance of girls to a greater degree than boys (Littleton, 1998; medium-weight evidence).

4.4 Other impacts reported as explanatory factors of the effects of using ICT

The authors of several studies included comments on other impacts of computer-based assessment; in particular, motivation, meaning-making, strategic awareness of learning processes and self-regulation. Since none of the studies included quantitative evidence of these impacts, or offered systematic qualitative evidence, the points made have therefore to be treated as tentative. They are included here because they point to some possible explanations of the impact of the computer on students' achievement.

An aim of Jackson's (1989) research was to find if computer-based testing might improve pupil motivation during a test by giving instant feedback and marking. In a study providing medium-weight evidence, Jackson reported that the students in the computer group 'were unanimous in their approval' and that 'the motivating factor of the computer played an important part' (p 814). However, the date of this study may be relevant here. In 1988, students were far less familiar with using a computer than is the case now, when perhaps the motivation derived from the novelty of using the computer would be much less. This would not, though, apply to the impact of the feedback.

Kumar *et al.* (1993) also provided medium-weight evidence that the immediate feedback provided by the computer program in their study of using a HyperCard-based approach to solving chemical equations motivated students to stay on-task until they had a solution that satisfied them.

Littleton *et al.* (1998) noted that the different performance of girls on problem-solving activities set in different contexts might be due to the way that characters in the scenario were made more realistic in the multimedia computer program. This may have encouraged girls to identify with the character, more so than boys, and so have influenced their motivation to engage with the task. This study provides medium-weight evidence.

The studies provide no definite findings about the impact of the assessment by computer on motivation but there are plausible arguments for motivation playing a part in improving performance on some computer-based tests.

Several studies used diagrammatic representations of events or ideas and relationships between them to support students' causal thinking and meaning-making. Masterman and Sharples (2002), in the context of causal thinking in history, suggested that the visual representation of causal relationships may have supported students' meaning-making of the events. However, Cifuentes and Yi-

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

48

Chuang (2000) found no improvement in meaning-making as a result of training middle school students in using computer graphics during study. They concluded that students needed more help in how to construct representations of concepts. Both these studies provide evidence of medium weight.

Studies involving concept/knowledge-mapping use visual display to provide a concrete form of abstract relationships. Kumar *et al.* (1993) expressed this as 'an added external memory' that reduces demand on working memory. O'Neil *et al.*, studying teamworking in concept-mapping, found some evidence (medium weight) suggesting that spending more time focusing on the map and less on reading messages from others, was associated with higher performance. They also pointed out that the task required knowledge that the students appeared to lack. When students do not possess understanding, it appears that this is likely to show in more complex tasks rather than less complex ones. Thus McFarlane *et al.* (2000) gave medium-weight evidence that lack of understanding was revealed in creating multimedia presentations that was not apparent in a conventional test.

Whilst there is some evidence from a study providing high-weight evidence that, through concept-mapping, students developed their understanding of relationships incrementally (Osmundson *et al.*, 1999), there is less clarity as to whether they were aware of the processes of their learning. The medium-weight evidence from the study by Masterman *et al.* (2002) found no examples of pupils spontaneously reflecting aloud on their reasoning skills.

In Tzuriel and Shamir's (2002) study of computer-mediated dynamic assessment, with students aged 4 to 7, the combination of teacher mediation with computer dynamic assessment was more effective that teacher mediation alone, the advantage being greater in more difficult tasks. This was high-weight evidence. It was suggested that a reason for this finding might be the feeling of competence derived from feedback from a human mediator being enhanced by feedback from the computer.

4.5 Quality assurance results

Data-extraction for all 12 studies was carried out independently by at least two people. In every case, one of these was one of the authors and the other person was either one of the four members of the Review Group who participated in this process (following training in using the EPPI-Reviewer) or the other author. EPPI-Centre staff also carried out a quality assurance role in the data-extraction process for a sample of studies.

On average, there were differences in 20 percent of questions. Most differences were of a trivial kind, such as in responding 'no', based on reviewer's judgement, as opposed to 'not stated' or 'implicit'. There were few differences in judgement, more in the degree of detail supplied to support judgements. The differences were readily reconciled by discussion, often with a third person. The details supplied by the different reviewers were combined in producing the final data-extraction for each study.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

49

5. FINDINGS AND IMPLICATIONS

After a brief overview of the findings reported in Chapters 3 and 4, this chapter discusses how the findings can be interpreted in terms of theoretical points foreshadowed in Chapter 1. It then identifies strengths and weaknesses of the review. Finally some implications for policy and practice, informed by the conference with users, are set out.

5.1 Identification and nature of studies

The search for studies was carried out through a process of handsearching journals online and in the library, searching relevant electronic databases, and using citations and personal contacts. It was a one-stage process, in which titles and abstracts were screened using the inclusion and exclusion criteria before being entered into the database. In this way a total of 103 studies were found. Full texts were obtained for 94 of these, from which 62 were excluded because they did not meet one of these criteria: concerned with assessment; ICT-based; concerned with students within the age range aged 4 to 18 in school; concerned with critical or creative thinking skills; or not reporting research. The remaining 32 studies were keyworded, a process that resulted in a further 20 being excluded. The characteristics of the 12 studies were used to map the identified research.

The 12 studies from which data were extracted were mainly identified by hand-searching journals or from electronic databases, four from the database of CRESST research. Seven studies were conducted by researchers in the US, although two of the CRESST studies were carried out with American students in Germany. Four studies originated in the UK and one in Israel. All used a computer program of some kind. Eleven studies were evaluations of situations set up by the researchers, in four of which students were randomly assigned to groups. A number of aspects of critical and creative thinking were involved in the assessments studied; only one assessed creative thinking as such. The largest number assessed the combination of critical and creative thinking through problem-solving, all of these being studies in secondary and middle schools. Studies in primary schools were disproportionately concerned with application of knowledge.

Eight studies reported findings in relation to impact of ICT assessment on students' achievement and three indicated that there was an impact on students' motivation (all in secondary or middle schools), although without reporting systematic data in support. Whilst half of the studies reported some impact on teachers, two specifically mentioned impact on teaching.

5.2 Summary of main findings

5.2.1 The findings in relation to the review questions

The main question for this review was:

- What evidence is there of the impact of the use of ICT for assessment of creative and critical thinking skills on students and teachers?

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

50

The identified studies provided information about three types of impact: on teachers' summative knowledge of students' achievements, on teachers' knowledge of students' processing and learning and on students' achievement. The combined findings relating to these three types of impact are summarised here.

Evidence from the high weight studies was as follows:

- Computer-based concept-mapping with automated scoring can be used to provide summative assessment of critical and creative thinking about complex relationships (Osmundson *et al.*, 1999).

- The use of ICT can help teachers by:
  - storing and recording information about how students are developing understanding of new material (Osmundson *et al.*, 1999);
  - taking over some of the role of assessing and providing feedback to students so that the teacher can focus on ways of supporting learning that are beyond the program's mediation (Osmundson *et al.*, 1999; Tzuriel and Shamir, 2002).

- Feedback from the computer during the use of test material improves student performance in later use of the same test material (Osmundson *et al.*, 1999; Tzuriel and Shamir, 2002).

The following findings were from studies providing medium-weight evidence in relation to the review question and should thus be treated more cautiously:

- Use of computers to assess teamwork did not provide evidence that aspects of collaboration result in increased problem-solving as measured by computer-based knowledge-mapping (O'Neil *et al.*,1999).
- Automated collection and scoring of the processes used in problem-solving provided additional information relevant to problem-solving performance (Schacter *et al.*, 1997; Schacter *et al.*, 1999).
- Using a computer program both to test and to give feedback to students increased the level of performance as compared with students taking the same tests on paper (Jackson, 1989).
- Using a computer-based program for solving problems of balancing chemical equations enabled students to achieve a higher level of performance than solving equations on paper (Kumar *et al.*, 1993).
- A conventional multiple-choice test gave a false impression of understanding compared with the analysis of multimedia presentations on the same topic (McFarlane *et al.*, 2000).
- Subject matter used in computer assessment of problem-solving affected the outcome for girls more than boys (Littleton *et al.*, 1998).
- The performance of middle school students was not improved by training in the use of computer graphics aimed at helping them make meaning of new material (Cifuentes and Yi-Chuan, 2000).
- Use of a program involving diagrammatic representation provided useful information about students' causal reasoning thinking through analysis of their diagrams but not from the log files of their computer moves (Masterman and Sharples, 2002).

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

51

- There was conflicting evidence from two studies relating to the impact on performance of using the web to search for information (Schacter *et al.*, 1997; Schacter *et al.*, 1999).

The subsidiary questions were as follows:

- How does any impact vary with the experience of the students and the conditions of the assessment?
- How does any impact vary according to whether the purpose of the assessment is formative or summative?
- What are the implications for assessment policy and practice of these findings? (This is addressed in section 5.5.)

The findings in relation to the first two of these questions from studies providing high-weight evidence, supported by others providing medium-weight evidence, were as follows:

- Interacting with a computer provides feedback that supports better performance even if this only reflects the moves and links made by students (Kumar *et al.*, 1993; Jackson, 1989; Masterman and Sharples, 2002; Osmundson *et al.*, 1999; Schacter *et al.*, 1997; Schacter *et al.*, 1999).
- Computers provide information about processes in reaching a solution that gives additional feedback to students and teachers (Schacter *et al.*, 1997; Schacter *et al.*, 1999; Osmundson *et al.*).

From studies of medium weight:

- There was evidence that the choice of content influences the performance of girls to a greater degree than boys (Littleton, 1998).
- Working with a team at a distance on concept-mapping was found not to improve performance as measured by knowledge-mapping (O'Neil *et al.*, 1999).
- Computer-based assessment enabled students to achieve at a higher level than in equivalent paper-based tasks (Kumar *et al.*, 1993).
- Using technology probed students' understanding to a greater degree than conventional tests (McFarlane *et al.*, 2000).
- Students' experience with computers and attitude towards them can influence computer-based test performance (Means *et al.*, 2000, Littleton *et al.*, 1998).

5.3 Discussion of findings in relation to the literature

5.3.1 The assessment of creative and critical thinking

The purpose of this review was to consider the use of ICT to extend the range of attainments that can be assessed and to facilitate a formative as well as a summative function of assessment, rather than to use ICT to replicate current paper-based assessments. In particular the review focused on the assessment of critical thinking (defined in terms of the evaluation of arguments or propositions in relation to evidence, reasoning, drawing conclusions) and creative thinking (in terms of relating together principles, ideas, information and entities in new and original ways to generate new entities or ideas). These key aspects of higher level thinking (Adey, 2001) were chosen because of their value in preparing

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

52

future citizens for life in a rapidly changing society and for life-long learning and because of their centrality in the learning process itself.

One of the points upon which there is general agreement, both among the authors of the studies considered in this review and others in the field, is that there is a lack of assessment of higher order thinking (Means *et al.*, 2000; McFarlane *et al.*, 2000; Schacter *et al.*, 1999; Baxter and Glaser, 1997; Baker, 1994). Schacter *et al.* (1999) point out that, even when assessment of higher level skills such as problem-solving is attempted by standardised tests, these provide 'numerous decontextualised, unconnected, and disparate problems' (p 404) which seem trivial and inconsequential and so do not adequately assess students' problem-solving ability, or the creative and critical thinking that authentic problem-solving involves. The reasons for this situation, they suggest, are based in the established procedures of test developers and the expectation of the general public of what tests and assessment are about.

The richer environment for problem-solving involving the use of creative and critical thinking skills that the use of ICT provides, as compared with paper-based tests, inevitably means that situations set up for assessment closely resemble those set up for developing these thinking skills. Thus learning is likely to take place during assessment. In eight of the 12 studies selected in this review (see Table 4.5) the research was aimed at both helping understanding and/or skill development as well as identifying students' shortcomings in their understanding and cognitive skills. Thus the assessment was intended to have both a formative as well as a summative purpose. These roles were not distinguishable in the studies, a point which leaves uncertain the extent to which the assessment can be separated from learning when ICT contexts are used.

5.3.2 Affordances of ICT for assessment of creative and critical thinking

As noted earlier, the term 'affordances' is used within a theoretical framework that sees ICT as a set of cultural tools offering opportunities for action (Barnes 2000, McFarlane 2001, Wertsch 1998). Studies reviewed here have indicated the potential of computer technologies to provide opportunities for students to use creative and critical thinking skills and to provide information for teachers in ways that can be adapted to assess a wide range of content. McFarlane *et al.* (2000), for example, described multimedia authoring as having the potential

> 'to support higher order thinking skills and processes. In turn the very processes entailed in producing such multimedia materials/writing are likely to assist the accomplishment of conceptual and procedural activities: for example the definition of relationships; consideration of the appropriateness (in terms of content and design) of information for different readers and different purposes; and the use of argument to establish or review different positions. …consequently, a framework for analysing the learning exemplified in pupils' multimedia presentations could be developed as a means of diagnostic, formative, evaluative or standards-referenced assessment of types of learning that are not easily revealed in current standardised tests' (p 205).

McFarlane *et al.* (2000) also point out that multimedia presentation provides students with the freedom to choose what ideas to link and how to present them 'at a stage in their cognitive development where their written expression is not sufficiently sophisticated to describe complex networks of associated ideas'

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

53

(p 205). This suggests that an impact of assessing in this way may be to scaffold learning by supporting their expression in ways other than writing. In doing so, it affords students better opportunities to show both what they do know and can do as well as revealing what they do not understand or are unable to do and thus gives teachers and others more valid information about students' thinking.

It is the non-linearity and dynamic representation afforded by forms of ICT that facilitates the expression of critical and creative thinking through making links and creating structures that reflect connections among ideas. Concept/knowledge-mapping facilitates non-linear representation of thinking. The findings of several studies reviewed here support the view that the process of concept-mapping is one of creating and evaluating links among concepts.

In order for assessment to have a formative purpose – that is, to 'bridge the gap between testing and instruction' (Schacter *et al.*, 1997) – it is necessary to be able to report not only the students' final performance, but also what processes students need to improve in order to raise their performance. The collection of information about processes, even if feasible in a non-computer-based assessment, is immensely time-consuming and would not be a realistic approach to meeting the need for information with value for improving learning. The uses of computers makes this information available, in some cases instantly, so that it provides essential feedback for the learner and the teacher.   The evidence of the studies by Schacter *et al.* (1997, 1999) and Osmundson *et al.* (1999) shows that collection of such information becomes a realistic possibility with the use of computer-based concept-mapping in a networked environment.

Whilst the studies provided evidence of the value of using ICT for information to the teacher, there were equivocal findings as to the impact of the computer use *per se* on students' attainment. Jackson's (1989) findings would suggest that the computer use increased scores, when the same test was given on paper and on computer; on the other hand, Tzuriel and Shamir (2002) found only a small effect from the use of the computer only. The factors explaining such difference in findings are likely to be related to motivation, the role played by the computer and the familiarity of the students with the computer. Jackson's study was conducted at a time when computers were being introduced into school and before the home computer became commonplace.   Thus the novelty factor could well have been a source of motivation. More recently, Russell and Haney (2000) pointed out that, whilst computer literacy will inevitably be an advantage for computer-based assessment, there is also evidence that active computer-users underperform on paper-based tests. However, Tzuriel and Shamir's study was with young children and showed an impact on performance when the computer provided feedback. This adds to the evidence that it is not merely the use of a computer that impacts on performance, but the way in which the computer is used.

When the process of assessment itself begins to impact on performance, then teaching and assessment begin to coalesce. The NCET (1994) report of 81 studies leads to the identification of a number of benefits of using ICT for learning which apply equally when computers are used for assessment. These include factors relating to speed of processing, which support speed of learning; elements of motivation, such as confidence, autonomy, self-regulation and enthusiasm, which support concentration and effort; ease of making revisions and improved presentation which support quality of writing and other products; and information handling and organisation, which support understanding. All these are benefits of a general kind that impact on students as learners and lead to the

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

54

tentative conclusion that there is a formative function in computer-based assessment when the program is designed with these factors in mind and also to give direct feedback to the student. Self-assessment is implicit here, although the studies have not directly reported on it.

In summary, the selected studies have, among them, indicated the potential for the affordances identified in Figure 2.1 in the following ways:

Tracking and representation of learner processes over time and cumulatively
Analysis of computer concept-mapping suggests that students develop understanding incrementally, but there is a lack of evidence of students being aware of their learning processes (Osmundson, 1999; Masterman, 2002). These studies provide medium-weight evidence.

Simulation of real- life scenarios for decision-making and problem-solving
Studies showed that real-life problems can be presented and represented in ways that call upon creative and critical thinking (Means *et al.*, 2000; McFarlane *et al.*, 2000; Schacter *et al.*, 1997; Schacter *et al.*, 1999). These provide medium- and low- weight evidence.

Visual, dynamic, non-linear data collection, analysis and feedback
The visual representation afforded by the use of computer programs in several studies was suggested as supporting students meaning-making (Masterman and Sharples, 2002; Kumar, 1993; O'Neil *et al.*, 1999). These provide medium-weight evidence.

The concrete form of abstract relationships provided by computer programs, particularly in concept/knowledge-mapping, was considered to be a factor in improved performance (Osmundson, 1999; Schacter *et al.*, 1997; Schacter *et al.*, 1999). These provide high- and medium-weight evidence.

Assessment of learning outcomes, which are presumed to have required creative and critical thinking skills
There were several studies indicating that a range of products and processes can be used to assess creative and critical thinking, often through problem-solving. (Jackson, 1989; Kumar *et al.*, 1993; Littleton *et al.*, 1998; McFarlane *et al.*, 2000; Means *et al.*, 2000; O'Neil *et al.*, 1999; Osmundson, 1999; Schacter *et al.*, 1997; Schacter *et al.*, 1999). These provide high-, medium- and low-weight evidence.

Speed of processing and feedback
The use of ICT made practicable the collection and use of complex information about the moves and processes that learner use in tackling assessment tasks, thus enabling this information to be used in supporting learning (Osmundson, 1999; Schacter *et al.*, 1997; Schacter *et al.*, 1999). These provide high- and medium-weight evidence.

Memory extension
The use of a computer program was considered to relieve demand on working memory by providing an 'external memory' for the student (Kumar, 1993). This study provides medium-weight evidence.

Development of self-awareness of learning and motivation for learning
In several studies, authors reported information evidence of the use of computer programs resulting in improved motivation for the assessment task, either through

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

55

the use of the computer *per se* or through the feedback it provided (Masterman and Sharples, 2000; Jackson, 1989; Kumar, 1993; Littleton, 1998). These provide medium-weight evidence.

Development of competence and motivation for learning
The impact of human mediation may be increased by computer mediation due to the computer enhancing the feeling of competence given by the teacher mediation alone (Tzuriel and Shamir, 2002). This study provide high-weight evidence.

5.4 Strengths and limitations of the review

5.4.1 Strengths of the review

The strengths of this review lie in the fact that it is a systematic review of research into a newly emerging field, which begins to demarcate some of the issues and problems as well as the substantive potential of ICT to support assessment for learning. Whilst the findings may be construed as limited, the review does bring together some important aspects of three different fields: ICT, learning and assessment. This can be built upon and developed.

One of the strengths of the review methodology lies in the specification of the research question and the care spent in deciding where and how to search for studies. The search process was systematic, which meant covering as many sources as possible, including all relevant electronic databases and journals, and following up citations in earlier reviews and in studies obtained for this review. A particular challenge was the amount of material available on the worldwide web, which is rapidly expanding and is often described as 'grey' material because it may not be published through regular academic channels. No search can, of course, be fully comprehensive but the strength of this methodology was in recording details of the search strategies and the search process.

A further strength of the systematic review was in the quality-assurance processes, which entailed checking key decisions on inclusion and exclusion of studies, keywording and data-extraction with two, and sometimes three, researchers. Conflicting decisions were readily resolved through discussion and negotiation between researchers.

The participation of peers in the review process was a particular strength. This was through the Review Group itself and through the consultation seminar at which the findings of the study were presented to experts in ICT, learning and assessment. Both these made a substantial contribution to the quality of the study through discussion and critique.

From the studies in this review, it is clear that various forms of ICT are used within particular educational paradigms, with their associated theories of learning. It may be particularly important that the influence of theory on the development of ICT be made explicit and a focus for research.

5.4.2 Limitations

One of the most significant features of this review is that the use of ICT for the assessment of creative and critical thinking skills is an emerging field, with a

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

56

small research base and varied practice. Very often, research reports that were found were actually accounts of the development of a methodology or of a software tool, rather than a report the impact of their use. The development of software is frequently driven by commercial or pragmatic interests, which may militate against sustained and more objective research.

As the review progressed, it became clear that the findings were relatively weak, and this may be accounted for by the complexity of the field and the small research base. Many of the findings relate to only one study and four of the final studies were associated with one institution. The review question presupposed a mature field of research and practice, and was in fact a complex question requiring expert knowledge of ICT, assessment practice and of higher order thinking skills. In fact, a prior question ('What are the affordances of ICT for the assessment of creative and critical thinking skills?') would have been more appropriate to the field and current knowledge. The outcomes as they stand in this review may not help researchers or practitioners because they are specific to particular programmes, and are difficult to generalise from in relation to impacts. However, looking at affordances allows a generalisation at a meta-level. Referring back to the research synthesis and identifying the key affordances of ICT as they are described in this review may be particularly useful as a checklist for practitioners, developers and researchers.

A further feature of this review, which proved challenging, was the blurring of boundaries between concepts previously assumed to be more clearly demarcated: for example, the boundaries between research, development and enterprise, between assessment and learning, between learning processes and learning outcomes, and between learner and ICT tool. Eight of the studies explicitly set out to foster as well as to assess creative and critical thinking skills; in other words, assessment was being used formatively and summatively. In these studies it did not prove possible to make a clear distinction between the aspects which were designed to explore the impact on pupils' achievement and those designed to assess that achievement.

The conceptual framework for synthesising and reporting the findings developed over time and was derived from the studies identified rather than a more widely based review of the potential of using ICT for assessment. Given the small number of studies, this suggests caution in the interpretation of the findings.

5.5 Implications of the review

The review procedures included a half-day consultation conference with assessment and ICT 'users' to discuss the findings of the review and the implications for policy, practice and research. The 20 participants were drawn from staff and higher degree students of the university who are involved in research, practice or study of information technology in education. They were asked initially to respond to a summary of the findings from the perspective of their experience and then, in groups, to consider implications.

In the initial discussion, participants indicated that, given the nature of the review questions and the emerging field of ICT use in assessment, the small number of studies identified was no surprise. The unclear distinction between development and research in many of the studies was also not unexpected and meant that several studies gave more information about the nature of affordances than empirical data about of impact. There was a strong feeling that the wide variety of

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

57

entities embraced by the term ICT made it impossible to generalise across studies and findings were specific to particular studies and technologies.

However, in discussing implications, it was considered that it was possible to go beyond the findings of individual studies using the participants' general knowledge of the field. The implications set out below draw in part on the group discussions.

5.5.1 Implications for policy

- The term ICT covers a range of diverse programs, applications, hardware and software, with varied affordances for assessment and for learning. It is important to recognise this and not to treat ICT as if it were a single entity.

- The use of any form of ICT in education should start from consideration of how it can be integrated into learning, teaching and assessment.

- Priority has to be given to conceptualising learning as having higher order thinking as a key feature for developing and assessing creative and critical thinking. Knowing how to use ICT should go hand in hand with understanding how to bring about higher level thinking.

- There is a need for professional development of teachers, teacher educators and advisers in the role that ICT in its various forms can take in learning, teaching and assessing higher order thinking.

- The route to effective use of ICT for learning, teaching and assessment in the classroom is through the teachers. Thus, it is important for teachers to recognise the affordances of ICT at first hand. Familiarity with using forms of ICT in their work and professional development to create and develop their own knowledge and understanding is a precursor to effective use of ICT in the classroom.

- Teachers should be provided with ways of evaluating software for its potential in developing and assessing critical and creative thinking, and opening up aspects of the subject not available elsewhere.

5.5.2 Implications for practice

- Learners need to be made aware of creative and critical thinking as explicit learning objectives so that they can use the feedback provided by ICT for formative self-assessment.

- Teachers should recognise their own role in exploiting the potential of ICT to provide intrinsic motivation for learning, through facilitating self-regulation, allowing learners more control of their learning and focusing their own role on aspects of mediation that are beyond the computer.

- The role of ICT in facilitating reflection on learning processes – the development of metacognition – should be made explicit to learners.

- Teachers need to be able to review software critically in relation to its potential for providing information for formative and summative assessment of

higher level thinking.

- Teachers need to be clear about the relationship/distinction between the roles that ICT can take in learning and assessment.

5.5.3 Implications for research

- The small number of studies identified, with a quarter of these emanating from one institution, is an indication of the need for a greater range of research studies on impact at the school level.

- There is need also for studies with different foci, in particular on the impact of specific affordances, such as non-linear representations, or feedback to teachers and students.

- More general studies of the affordances of ICT are needed, providing qualitative data as well as quantitative data that facilitate understanding of whether and if so, how, critical and creative thinking can be developed and assessed.

- The review raises conceptual and practical issues about the relationships between assessment and teaching and learning when ICT is used, which need to be addressed by researchers.

- In the view of participants at the consultation conference, it is important that all research is reported, and not suppressed on account of no impact or because of commercial interests.

6. REFERENCES

6.1 Included studies

Cifuentes L, Yi-Chuan J (2000) Concept learning through image processing. Paper presented at the 23rd National Convention of the Association for Educational Communications and Technology. Denver, CO, USA: October 25-28.

Jackson B (1988) A comparison between computer-based and traditional assessment tests, and their effects on pupil learning and scoring. *School Science Review* **69:** 809-815.

Kumar DD, White AL, Helgeson SL (1993) Effect of HyperCard and traditional performance assessment methods on expert-novice chemistry problem-solving. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching. Atlanta, GA, USA: April 15-19.

Littleton K, Light P, Joiner R, Messer D, Barnes P (1998) Gender, task scenarios and children's computer-based problem-solving. *Educational Psychology; an International Journal of Experimental Educational Psychology* **18:** 327-340.

Masterman L, Sharples M (2002) A theory-informed framework for designing software to support reasoning about causation in history. *Computers and Education* **38:** 165-185.

McFarlane A, Williams J, Bonnett M (2000) Assessment and multimedia authoring - a tool for externalising understanding. *Journal of Computer Assisted Learning* **16:** 201-212.

Means B, Penuel B, Quellmalz E (2000) Developing assessments for tomorrow's classrooms. In: *The Secretary's Conference on Educational Technology, 2000: Measuring Impacts and Shaping the Future.* Washington, DC, USA: US Department of Education.

O'Neil H, Chung G, Herl H (1999) *Computer-Based Collaborative Knowledge-mapping to Measure Team Processes and Team Outcomes.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California.

Osmundson E, Chung GKWK, Herl HE, Klein DCD (1999) *Knowledge-mapping In The Classroom: A Tool For Examining The Development Of Students' Conceptual Understandings.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://www.cse.ucla.edu/Reports/TECH507.pdf.

Schacter J, Herl HE, Chung GKWK, O'Neil Jr HF, Dennis RA, Lee JJ (1997) Feasibility of a web-based assessment of problem-solving. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL, USA: March 24-28.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

60

Schacter J, Herla H, Chung G, Dennis R, O'Neil H (1999) Computer-based performance assessments: a solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behaviour* **15:** 403-418.

Tzuriel D, Shamir A (2002) The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning* **18:** 21-32.

6.2 Studies found in initial search but not included in data-extraction

Agapova O, Ushakov A (1999) How technology changes education. *Technos* **8:** 27-31.

Aleven V, Koedingera K (2002) An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science* **26:** 147-179.

Appell L (1990) *Using Simulation Technology to Promote Social Competence of Handicapped Students. Final Report. Executive Summary.* Silver Spring, MD, USA: Macro Systems, Inc.

Bass G, Perkins H (1985) Teaching critical thinking skills with CAI: a design by two researchers shows computers can make a difference. *Electronic Learning* **4:** 32-34.

Benzie D (1999) Formative evaluation: can models help us to shape innovative programmes? *Education and Information Technologies* **4:** 249-260.

Berliner D (1993) Redesigning classroom activities for the future. *Educational Technology* **32:** 7-13.

Berns RG, Bailey DL (2001) *School-to-Work Professional Development Package for College and University Arts and Sciences Faculty, 2001.* Ohio, USA: School-to-Work Integration Coalition.

Beyer FS (1992) Impact of computers on middle-level student writing skills. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, USA: April 20-24.

Buchanan T (2000) The efficacy of a worldwide web mediated formative assessment. *Journal of Computer Assisted Learning* **16:** 193-200.

Buckley B (2001) Interactive multimedia and model-based learning in biology. *International Journal of Science Education* **22:** 895-935.

Chang C (2001) Construction and evaluation of a web-based learning portfolio system: an electronic assessment tool. *Innovations in Education and Teaching International* **38:** 144-155.

Chen GD, Liu CC, Ou K-L, Lin MS (2001) Web learning portfolios: a tool for supporting performance awareness. *Innovations in Education and Teaching International* **38:** 19-30.

Dimitriadi Y (2000) Evaluating the use of multimedia authoring with dyslexic learners: a case study. *British Journal of Educational Technology,* **32:** 265-275.

Divine K, Whanger R (1990) Use of a computer learning laboratory with at-risk high school students. *Educational Technology* **30:** 46-48.

Du Plessis JP, Van Biljona JA, Tolmiea CJ, Wollingera T (1995) A model for intelligent computer-aided education systems. *Computers and Education* **24:** 89-106.

ERIC (Educational Resources Information Centre) (1993) *Alternative Assessment and Technology. ERIC Digest ED365312.* Syracuse, NY, USA.

Eysink T, Dijkstra S, Kuper J (2002) The role of guidance in computer-based problem-solving for the development of concepts of logic. *Instructional Science* **30:** 307-333.

Fasick E, McLaren C (1995) The electronic student portfolio. *Kamehameha Journal of Education* **6:** 83-90.

Fenton R (1996) AMTC: Assessing motivation to communicate in the ninth grade speech class. Paper presented at the 82nd Annual Meeting of the Speech Communication Association. San Diego, CA, USA: November 23-26.

Fitch S (1993) *Outcome Based Alternative Assessment. Project SEED.* Auburn, ME, USA: Maine Center for Educational Services.

Fogarty R, McTighe J (1993) Educating teachers for higher order thinking; the three story intellect. *Theory into Practice* **32:** 161-169.

Frear V, Hirschbuhl J (1999) Does interactive multimedia promote achievement and higher level thinking skills for today's science students? *British Journal of Educational Technology* **30:** 323-329.

Gronlund WR, Redig KA, Sanders WA (1993) *Realistic Exercises for Assessing Learning. I. Conceptual Design and Testing. Report 04-93.* New London, CT, USA: Coast Guard Academy, Center for Advanced Studies.

Hague A, Benest I (1996) Towards over-the-shoulder guidance following a traditional learning metaphor. *Computers and Education* **26:** 61-70.

Hedberg JG, Brown C, Corderoy R (1994) Information landscapes and exploratory user interfaces: redesigning to improve learning outcomes. In: Simonson MR, Maushak N, Abu-Omar K (eds) *16th Annual Proceedings of Selected Research and Development Presentations*. Ames, Iowa: Iowa State University.

Higgins S (2001) Identifying feedback in mathematical activities using ICT. *Education* **29:** 3-13.

Hunter B, Bagley C, Bagley R (1993) Technology in the classroom: preparing students for the information age. *Schools in the Middle* **2:** 3-6.

Hurst K, Casillas A, Stevens R (1997) *Exploring the Dynamics of Complex Problem-Solving With Artificial Neural Network-Based Assessment Systems.* Los

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

62

Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://www.cse.ucla.edu/CRESST/Reports/TECH444.pdf

Jackson DF (1991) Exploratory problem-solving with a graphing application: semi-automated analysis of student performance data. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching. Lake Geneva, WI, USA: April 7-10.

Kajander A (1990) Measuring mathematical aptitude in exploratory computer environments. *Roeper Review* **12:** 254-256.

Kelly R (1999) Getting everybody involved: cooperative PowerPoint creations benefit inclusion students. *Learning and Leading with Technology* **27:** 10-14.

Klein D, O'Neil H, Baker E (1996) *A Cognitive Demands Analysis of Innovative Technologies.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://www.cse.ucla.edu/CRESST/Reports/TECH454.PDF

Klein D, O'Neil Jr H, Dennis R, Baker E (1997) The five families of cognitive learning: a context in which to conduct cognitive demands analyses of innovative technologies. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL, USA: March 24-28.

Klein D, Yarnall L, Glaubke C (2001) *Using Technology to Assess Students' Web Expertise.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California.

Koenig M (2001) Debating real-world issues. *Science Scope* **24:** 18-23.

Kurhila J, Laine T (2000) Individualized special education with cognitive skill assessment. *British Journal of Educational Technology* **31:** 163-170.

Lajoie S, Guerrera C, Munsie S, Lavigne N (2001) Constructing knowledge in the context of BioWorld. *Instructional Science* **29:** 155-186.

Lankes A (1996) *Electronic Portfolios: A New Idea in Assessment. ERIC Digest ED390377.* Syracuse, NY, USA: ERIC (Educational Resources Information Centre).

Lawrence L, O'Donnell A, Pallrand G (1994) Assessments in science: the development and evaluation of a video assessment in physics. *Educational Assessment* **2:** 275-293.

Levine T, Donitsa-Schmidt S, Zellermayer M (1996) Student perceptions of classroom climate in a communicative and computer supported approach to writing instruction. *Journal of Research and Development in Education* **29:** 94-103.

Lewis A, Smith D (1993) Defining higher order thinking. *Theory into Practice* **32:** 131-137.

Lin S, Liu E, Yuan S (2001) Web-based peer-assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning* **17:** 420-432.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

63

Lindsay M (1999) Designing assessment tasks to accomodate students cognitive skills in a technology-based mathematics course. *International Journal of Mathematical Education in Science and Technology* **30:** 691-697.

Linn M (1984) Policy issues in computer education. Assessing the Cognitive Consequences of Computer Environments for Learning (ACCCEL). Paper presented at the Annual Meeting of the American Educational Research Association. 68th, New Orleans, LA, USA: April 23-27.

Lipman M (1988) Critical thinking - what can it be? *Educational Leadership* **46:** 38-43.

Liu E, Lin J, Yuan S (2001) Design of a networked portfolio system. *British Journal of Educational Technology* **32:** 492-494.

Lore R, Chamberlain E (1990) *Language Development Component Secondary Reading Program, 1989-90. Final Evaluation Report*. Columbus, OH, USA: Columbus Public Schools.

Lynch L, Fawcett A, Nicolson R (2000) Computer-assisted reading intervention in a secondary school: an evaluation study. *British Journal of Educational Technology* **31:** 333-348.

McArthur D, Stasz C (1990) *An Intelligent Tutor for Basic Algebra*. Santa Monica, CA, USA: Rand Corporation.

McConnell D (1999) Examining a collaborative assessment process in networked lifelong learning. *Journal of Computer Assisted Learning* **15:** 232-243.

McFarlane A (2001) Perspectives on the relationship between ICT and assessment. *Journal of Computer Assisted Learning* **17:** 227-235.

Marshall G (1999) Exploring assessment. *Education and Information Technologies* **4:** 311-327.

Martens R, Hermans H (2000) Internet-based formative prior knowledge assessment. *Studies In Educational Evaluation* **26:** 245-258.

Mayer V, Raudebaugh W (1987) Pilot of a system for collecting daily classroom data on learning by using microcomputers. Paper presented at the Annual Meeting of the American Educational Research Association. Washington, DC, USA: April 20-24.

Merrimack Education Center (1986) *Technology Applications in Basic Skills (TABS). Report of Findings --1985. Second Year Summary.* Chelmsford, MA, USA.

Mislevy R, Steinberg L, Almond R, Haertel G, Penuel W (2001) *Leverage Points for Improving Educational Assessment.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://www.cse.ucla.edu/Reports/newTR534.pdf.

Mislevy R, Almond R (1997) *Graphical Models and Computerized Adaptive Testing.* Los Angeles, CA, USA: National Center for Research on Evaluation,

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

64

Standards and Sudent Testing, University of California. Available from: http://www.cse.ucla.edu/Reports/TECH434.PDF.

Mislevy R, Steinberg L, Breyer G, Almond R, Johnson L (1998) *A Cognitive Task Analysis, with Implications for Designing A Simulation-Based Performance Assessment.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://www.cse.ucla.edu/Reports/TECH487.PDF.

National School Boards Association (1997) T*echnology for Students with Disabilities: A Decision Maker's Resource Guide.* Alexandria, VA, USA.

O'Neil Jnr H, Klein D (1997) *Feasibility of Machine Scoring of Concept Maps.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://cresst96.cse.ucla.edu/Reports/TECH460.pdf.

O'Neil Jnr H, Schacter J (1997) *Test Specifications for Problem-Solving Assessment.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://cresst96.cse.ucla.edu/Reports/TECH463.pdf.

O'Neil Jnr H, Chung G, Brown R (1995) *Measurement of Teamwork Processes Using Computer Simulation.* Los Angeles, CA, USA: National Center for Research on Evaluation, Standards and Sudent Testing, University of California. Available from: http://www.cse.ucla.edu/Reports/TECH399.PDF.

Panagiotakopoulos C, Ioannidis G (2002) Assessing children's understanding of basic time concepts through multimedia software. *Computers and Education* **38:** 331-349.

Pappas P, Pearson J (1999) Take a skills snapshot: employing online self-assessments. *Multimedia Schools* **6:** 16-19.

Parker M (2000) Web-based extended learning through discussion forums*.* Research paper: connecting technology to teaching and learning. Paper presented at Connecting @ the Crossroads, National Educational Computing Conference. Atlanta, GA: June 26-28.

Peterson N, Sellers D (1992) Student motivation and learning styles in a multimedia learning environment. Paper presented at the Annual Meeting of the Northern Rocky Mountain Educational Research Association. Custer, SD, USA: October 9

Powers D, O'Neil K (1993) Inexperienced and anxious computer users: coping with a computer-administered test of academic skills. Edu*cational Assessment* **1:** 153-173.

Rennie LJ, Treagust DF, Kinnear A (1992) An evaluation of curriculum materials for teaching technology as a design process. *Research in Science and Technological Education* **10:** 203-217.

Rieth H (1990) C*urriculum Issues In Secondary School Programs For Students With Mild Disabilities*. Washington, DC, USA: COSMOS Corporation.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

65

Ross S, Casey J (1994) Using interactive software to develop students' problem-solving skills: evaluation of the 'intelligent physics tutor'. Paper presented at the 16th National Convention of the Association for Educational Communications and Technology. Nashville, TN, USA: February 16-20.

Russell M, Haney W (2000) Bridging the gap between testing and technology in schools. *Education Policy Analysis Archives* **8:** Article 19.

Safrit M, Ennis C, Nagle F (1988) The use of problem-solving skills in computer-aided instruction: an evaluation. *Journal of Computer Assisted Learning* **4:** 227-243.

Sandberg J, Anjewierden A, Groothuismink J, de Hoog R, Giebels M (2000) The CREDIT project: towards a generic framework for assessment and accreditation - a prototype for childcare. *British Journal of Educational Technology,* **31:** 277-294.

Slator B, Juell P, McClean P, Saini-Eidukat B, Schwert D, White R, Hill C (1999) Virtual environments for education. *Journal of Network and Computer Applications* **22:** 161-174.

Sluijsmansa D, Moerkerkea G, van Merriënboera J, Dochy F (2001) Peer-assessment in problem-based learning. *Studies In Educational Evaluation* **27:** 153-173.

Smeets E, Mooij T (2001) Pupil-centred learning, ICT, and teacher behaviour: observations in educational practice. *British Journal of Educational Technology* **32:** 403-417.

Subhi T (1997) Who is gifted? A computerised identification procedure. *High Ability Studies* **8:** 189-211.

Swaak J, de Jong T (2001) Learner vs. system control in using online support for simulation-based discovery learning. *Learning Environments Research* **4:** 217-241.

Swaak J, de Jong T (1996) Measuring intuitive knowledge in science: the development of the what-if test. *Studies In Educational Evaluation* **22:** 341-362.

Thelwall M (2000) Computer-based assessment: a versatile educational tool. *Computers and Education* 34: 37-49.

Tierney R (1988) *The Engagement of Thinking Processes: A two-year study of selected Apple Classroom of Tomorrow students.* Columbus, OH, USA: Ohio State University.

Tittle C, Hecht D (1992) A technology-based innovation to facilitate teacher access to and use of assessments of JHS students' self-regulatory, affective and motivational beliefs about mathematics. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, USA: April 20-24.

Tobin K (1998) Qualitative perceptions of learning environments on the world wide web. *Learning Environments Research* **1:** 139-162.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

66

Tsai C, Liu E, Lin S, Yuan S (2001) A networked peer-assessment assessment system based on a Vee Heuristic. *Innovations in Education and Teaching International* **38:** 220-230.

Van der Linden E (1993) Does feedback enhance computer-assisted language learning? *Computers and Education* **21:** 61-65.

Watts M, Lloyd C, Jackson C (2001) ICT in the literacy hour: studying a classroom innovation. *Research in Education* **66:** 84-97.

Watts M, Lloyd C, Jackson C (2001) ICT in the Literacy Hour: studying a classroom innovation. *Research in Education* **66:** 84-97.

Wegerif R (1996) Using computers to help coach exploratory talk across the curriculum. *Computers and Education* **26:** 51-60.

Zucker A, Esty E (1993) Promoting discourse in mathematics classrooms using a new video series for middle schools. Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta, GA, USA: April 12-16.

Zumbach J, Reimann P (2002) Enhancing learning from hypertext by inducing a goal orientation: comparing different approaches. *Instructional Science* **30:** 243-267.

6.3 Other references used in the text of the report

Adey PA (2001) The categorisation and assessment of 'thinking skills'. Paper presented at the EARLI conference. Fribourg, Switzerland: August 28-September 1.

Baker EL (1994) Learning-based assessment of history understanding*. Educational Psychologist* **29**: 97-106.

Barnes S (2000) What does electronic conferencing afford distance education? *Distance Education* **21**: 236-247.

Baxter G, Glaser R (1997) *An Approach to Analysing the Cognitive Complexity of Science Performance Assessments*. CSE Technical Report 452. Los Angeles, USA: University of California, Center for the Study of Evaluation.

Black P, Harrison C, Lee C, Marshall B, Wiliam D (2002) *Working Inside the Black Box.* London: King's College London, School of Education.

Black PJ, Wiliam D (1998) Assessment and classroom learning. *Assessment in Education* **5**: 7-73.

Brown A, Palinscar AS (1982) Inducing strategic learning from text by means of informed self-control thinking. *Topics in Learning Disabilities* **2**: 1-17.

Bunderson CV, Inouye DK, Olsen JB (1988) The four generations of computerized educational measurement. In: Linn RL (ed) *Educational Measurement.* 3rd edition. New York: Macmillan, pages 367-408.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

67

Claxton G (1999) *Wise-Up. The Challenge of Lifelong Learning.* New York, USA: Bloomsbury Publishing.

Crooks T (1988) The impact of classroom evaluation practices on students. *Review of Educational Research* **58**: 438-481.

Cuban L (1984) Policy and research dilemmas in teaching of reasoning: unplanned designs. *Review of Educational Research* **54**: 655-681.

De Bono E (1983) The direct teaching of thinking as a skill. *Phi Delta Kappa* **64**: 703-708.

Ennis RH (1993) Critical thinking assessment. *Theory into Practice*, **32**: 179-186.

Ennis R H (1996) Critical thinking dispositions: their nature and assessability. *Informal Logic* **18**: 165-182.

EPPI-Centre (2002) *Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research (version 0.9.5).* London: EPPI-Centre, Social Science Research Unit.

Fogarty R (1992) Beyond test scores: tracking significant outcomes. *Cogitare* **6**: 2.

Fogarty R, McTighe J (1993) Educating teachers for higher order thinking: the three-story intellect. *Theory into Practice* **32**: 161-169.

Gordon S, Reese M (1997) High stakes testing: worth the price? *Journal of School Leadership* **7**: 345-368.

Harlen W, Deakin Crick R (2002) A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Huitt W (1998) *Critical Thinking.* Available on: http://chiron.valdosta.edu/whuitt/col/cogsys/critthnk.html.

Kellaghan T, Madaus G, Raczek A (1996) *The Use of External Examinations to Improve Student Motivation.* Washington DC, US: AERA.

Klieme E (2000) *Assessment of Cross-Curricular Problem-Solving Competencies*. Berlin, Germany: Max-Planck Institute for Human Development, Centre for Educational Research.

Lawson AE (1993) At what levels of education is the teaching of thinking effective? *Theory into Practice* **32**: 170-178.

Lawson AE, Abraham MR, Renner JW (1989) *A Theory of Instruction: using learning cycle to teach science concepts and thinking skills.* Cincinnati, OH, US: National Association for Research in Science Teaching.

Lewis A, Smith D (1993) Defining higher order thinking. *Theory into Practice* **32**: 131-137.

Linn RL (1994) Performance assessment: policy, promises and technical measurement standards. *Educational Researcher* **24**: 4-13.

Linn RL, Dunbar SB, Harnisch DL, Hastings CN (1982). The validity of the Title I evaluation and reporting system. In: House ER, Mathison S, Pearsol J, Preskill H (eds) *Evaluation Studies Review Annual*. Volume 7. Beverly Hills, CA, USA: Sage Publications, pages 427-442.

Marzano R, Hutchins R (1985) *Thinking Skills: A conceptual framework. A special issue of Noteworthy.* Aurora, CO, USA: Mid-continent Regional Educational Laboratory (ERIC document 266 436).

McFarlane A (2001) Perspectives on the relationship between ICT and assessment. *Journal of Computer Assisted Learning* **17**: 227-234.

McFarlane A (2002) Educating the inheritors of the information age. Inaugural lecture. Graduate School of Education, University of Bristol: November.

McFarlane A, De Rijke FJM (1999) Educational use of ICT. OECD Quality Assurance working paper for the Educational Software Working Group. Available from: http://waldorf.eds.udel.edu/oecd/quality/papers/papershome.html.

McGuiness C (1999) *From Thinking Skills to Thinking Classrooms: A Review and Evaluation of Approaches for Developing Pupils' Thinking*. DfEE Research Paper RR115. Norwich: DfEE.

National Council for Educational Technology (NCET) (1994) *Integrated Learning Systems: A Report of the Pilot Evaluation of ILS in the UK.* Coventry: NCET.

Newman FM (1990) Higher order thinking in teaching social studies: a rationale for the assessment of classroom thoughtfulness. *Journal of Curriculum Studies* **22**: 41-56.

OECD (1999) *Measuring Student Knowledge and Skills*. Paris: OECD.

Russell M, Haney W (2000) Bridging the gap between testing and technology in schools. *Educational Policy Analysis Archives* **8:** 19, ISSN 1068-2341.

Shepard L (1991) Will national tests improve student learning? *Phi Delta Kappa* **73:** 232-238.

Wertsch J (1998) *Mind as Action*. USA: Oxford University Press.

Yeh SS (2001) Tests worth teaching to: constructing state-mandated tests that emphasize critical thinking. *Educational Researcher* **30**: 2-11.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

69

APPENDIX 2.1: Search strategy

Key terms

ICT	Assessment	Creative and critical thinking skills	Relevance to school
Computer Technology Processor PC Lap top ICT Information & Communications Technology Information & Computer Technology World wide web Web based	Testing Assessment Profile Evaluation Self assessment Peer-assessment assessment Dynamic assessment Portfolio Appraisal Narrative Story Computer-based assessment	Thinking skills Problem-solving Cognitive skills Emotional literacy Affective skills Knowledge- mapping Mind mapping Critical thinking Values education Personal development Speculative thinking Creative thinking Emotional intelligence Creativity Deep learning Meaning-making Higher order skills Curiosity Metacognition	School Infant school Primary school Elementary school Secondary school Community school Urban school Suburban school Private school State school High school Middle school Pre-school Kindergarten

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

70

Search strategy for electronic databases

1   computer uses in education/ or Education/ or Educational technology/ or Information technology/ or Instruction/ or Instructional development/ or Instructional improvement/ or Instructional innovation/ or Instructional materials/ or Multimedia instruction/ or Programmed instruction/

2   Internet/

3   exp computer assisted instruction/ or exp distance education/ or exp educational technology/ or exp information technology/

4   exp curriculum evaluation/ or exp evaluation/ or exp formative evaluation/ or exp holistic evaluation/ or exp instructional material evaluation/ or exp peer evaluation/ or exp psychological evaluation/ or exp 'self evaluation (groups)'/ or exp 'self evaluation (individuals)'/ or exp speech evaluation/ or exp student evaluation/ or exp summative evaluation/ or exp teacher evaluation/ or exp vocational evaluation/ or exp writing evaluation/

5   exp evaluation/

6   Abstract reasoning/ or Basic skills/ or Cognitive development/ or Cognitive processes/ or Cognitive style/ or Communication skills/ or Comprehension/ or Concept formation/ or Convergent thinking/ or Creative thinking/ or Critical thinking/ or Decision-making/ or Decision making skills/ or Deduction/ or Divergent thinking/ or Evaluative thinking/ or Induction/ or Interpretive skills/ or Intuition/ or Language skills/ or Learning modalities/ or Learning processes/ or Learning strategies/ or Listening skills/ or Logical thinking/ or Mathematics skills/ or Metacognition/ or Problem-solving/ or Productive thinking/ or Reading skills/ or Research skills/ or Science process skills/ or Skill development/ or Study skills/ or Thinking skills/ or Visualisation/ or Writing skills/

7   exp affective behavior/ or exp emotional adjustment/ or exp emotional development/ or exp empathy/ or exp interpersonal competence/ or exp leadership/

8   exp emotional development/ or exp individual development/ or exp spirituality/ or exp values/

9   exp prior learning/ or exp reading strategies/ or exp student reaction/

10 exp british infant schools/ or exp catholic schools/ or exp charter schools/ or exp community schools/ or exp consolidated schools/ or exp day schools/ or exp disadvantaged schools/ or exp elementary schools/ or exp experimental schools/ or exp free schools/ or exp high schools/ or exp junior high schools/ or exp middle schools/ or exp private schools/ or exp schools/ or exp secondary schools/ or exp state schools/ or exp suburban schools/ or exp urban schools/

11 (1 or 2 or 3) and (4 or 5) and (6 or 7 or 8 or 9 or 10)

12 1 or 2 or 3

13 4 or 5

14 12 and 13 and 6 and 10

APPENDIX 2.2: Journals handsearched online or in library

Journal	Dates searched
American Educational Research Journal	92-02
American Journal of Education	92-02
American Journal of Evaluation	00-02
Assessment in Education	94-02
British Educational Research Journal	00-02
British Journal of Educational Psychology	99-02
British Journal of Educational Technology	98-02
Cambridge Journal of Education	00-02
Cognitive Science	94-02
Computers and Education	76-02
Contemporary Educational Psychology	93-02
Curriculum Journal	92-02
Developmental Review	93-02
Education 3-13	92-02
Education and Information Technologies	93-02
Education Review	92-02
Educational Assessment	92-02
Educational Technology	90-96
European Journal of Education	00-02
Harvard Education review	92-02
Innovations in Education and Training International	00-02
Instructional Science	97-02
Interactions (University of Warwick Educational Technology Service)	97-02
International Journal of Educational Research	86-02
International Journal of Mathematical Education in Science and Technology	99-02
Journal of Computer Assisted Learning	92-02
Journal of Educational Measurement	90-02
Journal of Information Technology for Teacher Education	93-02
Journal of Research and Development in Education	90-02
Learning and Instruction	91-02
Learning Environments Research	98-02
Oxford Review of Education	98-02
Research in Education	92-02
Research Papers in Education	90-02
Review of Educational Research	92-02
Studies in Educational Evaluation	92-02
Teachers College Record	92-02

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

72

APPENDIX 2.3: EPPI-Centre keyword sheet including review-specific keywords

EPPI-CENTRE EDUCATIONAL KEYWORDING SHEET v0.9.5

1. Identification of report
Citation
Contact
Handsearch
Unknown
Electronic database
 (*Please specify.*) …………………………

2. Status
Published
In press
Unpublished

3. 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.*)
………………………………………….

4. Language (*Please specify.*)

……………………………………

5. In which country/countries was the study carried out? (*Please specify*)

……………………………………

6. 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.*)
………………………………………

6a 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.*)
………………………………………..

7. Programme name (*Please specify.*)

………………………………………..

8. 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.*)………………………

8a Age of learners (years)
0-4
5-10
11-16
17-20
21 and over

*8b. Sex of learners
Female only
Male only
Mixed sex

9. 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.*)…

10. 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 assessment information is provided?	What form of ICT is used?	Which creative and critical thinking skills are assessed?	What impact is reported?
Formative feedback to teacher Formative feedback to student Summative for teacher Summative for student Peer-assessment assessment	Multi-media Computer program World wide web Intranet/Network	Problem-solving Communication skills Access and assessing information Metacognitive reflection Application of knowledge Affective skills Generation of new ideas or products	On students' motivation On students' achievement On teachers' teaching On teachers' perception of students

APPENDIX 3.1: Tables of cross-tabulations of keywords

Table 3.1.1: Type of information provided in studies in different educational settings

Assessment information provided (categories are not exclusive)	Educational setting (N=12)		
	Nursery school (Total = 1)	Primary school (Total = 3)	Secondary school (including Middle) (Total = 8)
Formative feedback to teacher	0	3	3
Formative feedback to student	1	2	6
Summative for teacher	0	2	7
Summative for student	0	0	0
Peer-assessment assessment	0	0	0

Table 3.1.2: Form of ICT used in studies in different educational settings

Form of ICT used (categories are not exclusive)	Educational setting (N=12)		
	Nursery school (Total = 1)	Primary school (Total = 3)	Secondary school (including Middle) (Total = 8)
Multi-media	0	1	0
Computer program	1	3	8
World wide web	0	0	0
Intranet/Network	0	0	4

Table 3.1.3: Critical and creative thinking assessed in studies in different educational settings

Creative and critical thinking assessed (categories are not exclusive)	Educational setting (N=12)		
	Nursery school (Total = 1)	Primary school (Total = 3)	Secondary school Total = 8)
Problem-solving	0	0	5
Communication skills	0	0	3
Access and assessing information	0	1	2
Metacognitive reflection	0	1	1
Application of knowledge	1	2	2
Affective skills	0	0	2
Generation of new ideas or products	0	0	1

Table 3.1.4: Type of impact reported in studies in different educational settings

Type of impact reported (categories are not exclusive)	Educational setting (N=12)		
	Nursery school (Total = 1)	Primary school (Total = 3)	Secondary school (Total = 8)
On students' motivation	0	0	3
On students' achievement	1	2	5
On teachers' teaching	0	1	1
On teachers' perception of students	1	1	2

Table 3.1.5: Type of impact reported in studies providing different assessment information (Note that categories are not independent.)

Assessment information provided	Type of impact reported (N=12)			
	On students' motivation Total = 3	On students' achievement Total = 8	On teachers' teaching Total = 2	On teachers' perception of students Total = 4
Formative feedback to teacher	0	5	2	1
Formative feedback to student	2	8	2	1
Summative for teacher	3	8	1	0
Summative for students	0	0	0	0
Peer-assessment assessment	0	0	0	0

Table 3.1.6: Type of impact reported in studies assessing different creative and critical thinking skills (Note that categories are not independent.)

Creative and critical thinking assessed	Type of impact reported (N=12)			
	On students' motivation Total = 3	On students' achievement Total = 8	On teachers' teaching Total = 2	On teachers' perception of students Total = 4
Problem-solving	1	5	0	0
Communication skills	0	2	1	1
Access and assessing information	0	3	0	0
Metacognitive reflection	0	2	0	0
Application of knowledge	1	4	1	0
Affective skills	1	2	0	0
Generation of new ideas or products	0	0	1	1

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

Cifuentes L and Yi-Chuan J (2000) Concept learning through image-processing

Setting: Middle school   **Country:** USA

Broad aims
The study explored computer-based image processing as a study strategy for middle school science concept learning.

Topic focus
The study is concerned with how the manipulation of images during knowledge construction in science can support learners' meaning-making and conceptual development, and how the students' graphical representations can inform teacher feedback to students.
Experimental group participated in workshop of three 50-minute training sessions on 'how to manipulate and generate computer graphics during study time using materials developed by researchers in Apple Works, Photoshop, and Powerpoint. They then received the same print-based, verbal material that the control group received' (on a science concept - general properties of matter)... 'and students were given 50 minutes to study prior to a test. At the end of the 50 minutes students kept both their study notes and reading material. The next day, and prior to taking the test, students handed in their study notes' (p 77). Workshop had five objectives: (1) recognise underlying structure of text, (2) illustrate underlying structure, (3) relate new concepts to prior knowledge, (4) highlight distinctive features, (5) use graphics for review Data sources: (1) immediate recall test (multiple-choice paper and pencil test), (2) students' study notes, (3) students' computer files, (4) a web-based 'Study Strategies Questionnaire'

Research questions
What are the effects of computer graphics generation on science concept learning and the impact of using computer graphics to show interrelationships among concepts during study time?

Study design
87 students from 7th and 8th grade classes were randomly assigned to treatment group or a control group, each receiving different interventions relating to teaching science concepts. Post-test measures were administered to both groups for comparison. Post-test only control group design.

Data collection and analysis
The three 50-minute workshop sessions on how to manipulate and generate computer graphics during study time had the following objectives: for students to be able to (a) recognize underlying structure of text (inter-relationships), (b) illustrate underlying structure, (c) relate new concepts to prior knowledge, (d) highlight distinctive features, and (e) use graphics for review.

The researchers facilitated the workshop. They modelled visualisation of concepts using seven underlying structures, gave the students a turn with each of the seven underlying structures, modelled direct representation of concepts and highlighting distinctive features, gave the students a turn with direct representation and highlighting, and gave students 15 short texts to visualize. They encouraged students to keep their graphic representations and use them for study and review. After taking the immediate recall test, participants filled in a web-based 'Study Strategies Questionnaire' that asked them to rate the content that they had previously been exposed to the information in 'General Properties of Matter?' To determine if groups varied in their prior knowledge of the textual material, a t-test was conducted. No difference was found. The questionnaire also asked students to describe in detail the steps that they took to prepare for the test. The testing instrument for immediate recall contained 30 multiple-choice items. All students took the immediate recall rest at the end of their 50-minute study to determine the effects of the experimental treatment. In addition, all participants' study notes and printouts of computer graphics were collected. Participants were asked on the web-based survey to describe in detail the steps that they took to prepare for the test. The researchers rated the participants as visualisers or non-visualisers based upon the students' study notes and study strategies reported on the survey. Students were classified as visualisers if they used the computers to construct visuals while they studied for the rest and/or reported that they |

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

77

highlighted or drew while they studied. Students were classified as non-visualisers if they did not create visuals or highlight during study. The effects of treatment and application of visualisation during study time on immediate recall were then estimated by comparing scores using planned contrasts in a general linear model.

Results

The Analysis of Variance (ANOVA) revealed a negative treatment effect. The control group performed better than the experimental group and there was no statistical difference between the scores of the visualisers and non-visualisers (judged from whether the students used graphical representation in their notes). There was no interaction between group and strategy on students' performance. Learner generated graphic representations of concepts provide a rich resource for the students' teachers.' (p 78)

Conclusions

Learner-generated graphic representations of concepts provided a rich resource for students' teachers. Representations of learners' understandings provided teachers with a way of knowing whether or not students grasp concepts. Teachers suggested that, if students cannot visualize the concept, perhaps they do not thoroughly understand the concept. In summary, the research delivered a visualisation workshop designed to help learners use computer graphics to construct meaning while they study. When delivering visualisation workshops in the natural setting of schools, problems are bound to arise; technical failure, human interruptions, lack of active participation, limited access, distraction by alternate tools or games on computers, and differences in learners' abilities each affect the success of a workshop. Middle-school students are unsophisticated learners and require guidance toward effective visualisation. Given the three-day workshop, students were unable to internalise the visualisation methods as part of their study strategy. Most students were cognitively not ready to generate meaningful computer graphics while they study the textual information. They were more likely to highlight the important points by typing those words or sentences on computer than producing visuals that could help make the concepts concrete. Middle-schoolers might need extensive practice in constructing their own representations whilst receiving expert feedback regarding their appropriateness

The 7th and 8th grade students who participated in our workshop were not sophisticated visualisers. They needed expert modelling of identifying underlying structure of texts. It is essential to have all students engaged in the diverse practice of constructing their own concept representations whilst receiving expert feedback regarding their appropriateness.

Weight of evidence

A: Medium   B: Medium   C: Low   D: Medium

Jackson B (1989) A comparison between computer-based and traditional assessment tests, and their effects on pupil learning and scoring

Setting: Middle school   **Country:** England

Broad aims

'This study is concerned with the use of a microcomputer in the assessment of pupils in the sciences in a middle school, and its effect on their motivation, attitudes and learning, particularly with regard to providing immediate feedback to the pupil who gives incorrect responses.' (p 809)

Topic focus

The study sets out to discover if a testing programme based on a computer could give any significant educational advantage to the pupil. That is, could it improve pupil motivation during a test by giving instant feedback and marking, therefore improving understanding and giving an enhanced score in a future test? Two groups: control group ('pencil and paper') and experimental group ('computer group') were given a multiple-choice test of 40 items. During tests carried out on a computer, the pupil would see either a congratulatory message or a comment to allow for a second attempt - and then either congratulation or the correct response.

Research questions

Can a testing programme based on a computer give any significant educational advantage to the pupil? That is, could it improve pupil motivation during a test by giving instant feedback and marking, therefore improving understanding and giving an enhanced score in a future test?

Study design

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

78

The study was a randomised controlled trial, with two groups being selected from two fourth-year science classes. The experimental group undertook assessments using multiple-choice science tests on computers and the control group undertook assessments using pen and pencil.

The outcomes were compared in a final test which included all the previous questions addressed and the multiple-completion (cloze procedure) test.

Data collection and analysis

Data were collected though 40 multiple-choice questions selected from the author's item bank. These were divided into four sets. In addition, a cloze procedure test was administered in which students had to add keywords which were missing. Half of these were administered via a computer program with immediate feedback to students, which was in the form of either a congratulatory message if the question was correct or another comment related to the question if it was answered incorrectly. This would help the student in another attempt. If this second attempt was correct, the student was congratulated; if it was wrong, then the correct response was given before the student moved on.

Results

The overall mean score for the paper/pencil groups was lower than the computer group (26.70 and 30.96 respectively). There was greater distribution around the mean in the pencil/paper group. The t-test conducted on the overall scores of the two groups was significant at the 0.05 level. The final cloze tests all had a low score.

Conclusions

Pupils who are tested with multiple-choice questions when computers provide feedback and prompts are likely to achieve higher grades on later tests than those who receive no feedback through the traditional paper/pencil route.

Pupils engaged in the computer group were unanimous in their approval of the computer-based tests. The administrative gains were worthwhile; the speed of feedback and to look back at individual results was useful for formative assessment.

'Using the computer to both test and give immediate feedback to the pupil is only worthwhile if there is some evidence that it helps pupil understanding and learning. The evidence of this study indicates that at the 5.5% level of significance, pupils who have been tested in this way do score more highly in a later test using the same material, than those who are tested using the traditional method.' (p 814)

Weight of evidence

A: Medium   B: Medium   C: Low   D: Medium

Kumar D, White A, Helgeson S (1993) Effect of HyperCard and traditional performance assessment methods on expert-novice chemistry problem-solving

Setting: High school   **Country**: USA

Broad aims

To study the relative effects of a HyperCard-based approach to solving stoichiometric equations and traditional paper-based methods on the performance of expert and novice high school chemistry students.

Topic focus

This study examined the relative effect of a HyperCard-based approach to balancing chemical equations and a conventional paper-based approach. The progam registers all the students' moves including overall time to complete the test, provides immediate feedback to the students, and provides a record for the teacher. The students received relevant instruction in chemical equations about four weeks before the trial. The sample consisted of 30 'honours' students (experts) and 30 regular chemistry students (novices). These groups were divided into 15 each using HyperCard and paper versions of the equations.

Research questions

Null hypotheses:
(1) There is no significant difference between the HyperCard and traditional paper-based assessment methods for balancing chemical equations
(2) There is no significant difference between the performance of expert and novice students using the HyperCard and traditional paper-based methods for balancing chemical equations.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

79

Study design

The design was to compare two forms of the assessment for both students who were 'expert' in chemistry (used to solving problems for themselves and preparing for lab experiments by themselves) and those who were novices (had less experience of chemistry than the 'experts'). The novice and expert groups were randomly assigned to using either the HyperCard or paper-based method of balancing chemical equations.

Data collection and analysis

Data collected (all used to measure aspects that were findings of the study):
number of attempts at solving the equations;
total time on task;
correctness of the answer (1 or 0)

Results

Students with high performance scores correctly balanced more chemical equations than students with low performance scores.
Students with high performance scores required fewer attempts to balance the chemical equations than students with low performance scores.
Students with high performance scores had a lower rate of attempts than students with low performance scores.
Performance scores for both expert and novice students were significantly higher on the HyperCard method than on the paper-based method.
Performance scores of novices using the HyperCard methods nearly equalled the performance score of the experts using the paper-based method.
Number of equations correct for both expert and novice students were significantly higher on the HyperCard method than on the paper-based method.
Significant interactions were found for time on-task and number of correct equations.
Time on-task for the experts decreased from paper-based to HyperCard while time on-task for the novices increased from paper-based methods to HyperCard.
The slope of the methods of correctness plot for the novices from paper-based to HyperCard is greater than that for the experts, indicating that the novices benefited from the HyperCard more than the experts.

Conclusions

The performance of both groups was higher in using HyperCard for balancing equation than a paper-based method. 'It is possible that the non-linear environment of the HyperCard and the computer platform may have played a significant role in helping the novice problem solver improve performance by providing for thinking in a less restrictive way'.
The authors also suggested:
that the mouse interface was less interfering that using a pen in solving the equations;
that the HyperCard method of the test provided immediate feedback so that the students was motivated to stay on-task until a satisfactory solution was reached;
the computer itself provided an added external memory for the student while balancing the equations, thereby reducing the cognitive demand on working memory'.

Weight of evidence

A: Medium B: High C: Medium D: Medium

Littleton K *et al.* (1998) Gender, task scenarios and children's computer based problem-solving
Setting: Secondary school   **Country:** England
Broad aims
The extent to which superficial aspects of the software in use can affect the performance of girls and boys on computer-based, problem-solving tasks.
Topic focus
The study is concerned to verify the source of difference between girls' and boys' performances on problem-solving tasks. Similar problem-solving programs were created using difference subject matter. In 'King and Crown', the quest takes the form of the retrieval of a crown from an island. The 'Honeybear' task is designed to offer exactly the same formal task demands but within a different scenario. Both tasks present the same problem in the same adventure-game format and call for an identical solution strategy. There were two studies each involving equal numbers of boys and girls and comparing performance in the same way. In the second study, a different version of 'King and Crown' was designed, called Pirates. In each case, half the boys and half the girls were randomly assigned to one or other of the scenarios. Children were also given an attitude test immediately after the problem-solving.
Research questions
The study is concerned to verify the source of difference between girls' and boys' performances on problem-solving tasks. It addresses the question of how far superficial aspects of the software in use can affect the performance of girls and boys on computer-based problem-solving tasks.
Study design
The study compared the performance on two different problem-solving tasks of 26 boys and 26 girls in study 1 and 24 boys and 26 girls in study 2. The children were given instructions about using the computer and then tackled the task individually. There was a time limit of 30 minutes and the computer logged the moves of the children even if they did not solve the problem. The children were given a short paper and pencil attitude-rating questionnaire at the end of the task with items relating to their liking for computers in general and their enjoyment of the task.
Data collection and analysis
Data were collected about the number of children solving the problem and they were scored on progress towards the task solution. They were also given an attitude-rating questionnaire. Information was also collected about gender and type of problem.
Results
Study 1: The authors report that 'children's problem-solving performance was significantly affected by the scenario within which the problem was cast and, in particular, the performance of girls was strongly influenced by the version of the task employed'. Study 2: 'As in the previous study, the performance of the boys remained relatively unaffected by the software type, whereas the performance of the girls is far superior when using the 'Honeybears' version as compared with the Pirates version.' The results of this study replicate the software/gender interaction seen in study 1 and offer evidence that it is the difference in the scenario, as opposed to any confounding differences in the characteristics of the interface, which are accounting for these effects.
Conclusions
'Context exerts a critical influence on cognitive performance and can affect not just the absolute difficulty of a task, but also the relative difficulty of tasks for different groups of children. Educational software designers need to heed the lesson now widely learned by psychologists, namely, that vehicles for task presentation are rarely, if ever, neutral in their effects.' (p338) The performance of girls on computer-based problem-solving tasks is substantially affected by the nature and focus of the software used, whereas boys performance remains the same. Girls identified with the characters in the software and personified the characters in a way that was not evident amongst the boys. The authors suggest that this factor may have influenced the girls' motivation to engage with the task and may impact the level of confidence girls have with using computers, which may be more important than gender per se. Gender differences in response to computer programs can be substantial but they are far from immutable. The imagery or metaphors used in the presentation of the task can have an influence out of all proportion to their significance to the designer.

Weight of evidence
A: High B: Medium C: Low D: Medium

Masterman L and Sharples M (2002) A theory-informed framework for designing software to support reasoning about causation in history
Setting: High school   **Country:** USA
Broad aims
'To elicit usability issues that facilitate or, conversely, interfere with pupils' use of a computer program designed to support reasoning about causation in history. To ascertain how the program might be used to identify shortcomings in pupils' causal thinking.'
Topic focus
The study is a formative evaluation of a computer program designed to: 1. Stimulate pupils to refine their understanding of causation and its associated concepts through constructing and interacting with diagrammatic representations of the causal relationships of specific historical situations. 2. Make visible the capabilities and limitations of pupils' causal reasoning through (a) completed representation and (b) records of the actions taken by pupils in constructing their representations'. In the program, the pupils create diagrams linking 'cause boxes' (different causal factors) by arrows to the final event.. They start by creating a cause box and assigning properties to it (long or short term, religious, economic, etc.). When they have created more than one box, they draw arrows to denote relationships. The computer records a log of their moves. The initial theoretical framework was based on socio-cognitive and socio-constructivist models. This was developed through the use of the concept of the 'zone of proximal development' (ZPD), which includes both development potential in relationship, and the support structure and cultural tools in use. The computer program can be either pedagogically engaged or pedagogically neutral. The concepts of the modelling-coaching-fading strategy and the 'conversational metaphor are also utilised. These concepts informed the development of the software.
Research questions
Aims of program: to help pupils' understanding of causation in history Aims of evaluation: to elicit usability issues that facilitate, or conversely, interfere with pupils' use of the program to ascertain how the program might be used to identify shortcomings in pupils' causal thinking
Study design
The same intervention was provided to all pupils in four classes. Data were collected about the processes of using the computer program and about pupils' and teachers views on usability, easiness, etc.
Data collection and analysis
Data was collected by pupil questionnaire, and from the records made by the computer of the pupils' diagrams showing, the number of cause boxes, the number of links made between boxes, the ratio of long-term to short-tem causes and the ratio of causal links to non-causal links. Some diagrams were completed by pupils in pairs and these were treated in the same way as for individual pupils'. Informal discussion with teachers were also used as evidence.
Results
Findings relating to usability of the program: 'Over 92% of pupils reported that the program was easy or very easy to use. Log file showed few procedural errors. Nearly 90% considered it easier to draw a diagram with the program than using paper and pencil, and 78% were willing to use the program again in
another history topic'.' The quantitative analysis of the diagrams suggested that most pupils seemed able to distinguish between causes and facts. Of the diagrams, 84% contained links with the average of 69% of the boxes linked to at least one other box. However, less than one-third of links could be classified as causal in nature and the prevalence of fragmented chains indicated that pupils had great difficulty linking causes into any coherent structure. Inspection of the log files for evidence of pupils' patterns of behaviour produced just one useful example. The transcripts of pupils' presentations did not yield any example in which pupils spontaneously reflected

aloud on their reasoning skills.

Most pupils were able to distinguish between causes and mere facts. However, pupils had difficulty linking causes into any coherent structure; that is they judged each cause in isolation from the other causes.

The diagrams developed through the use of the computer program can serve as a powerful aid to articulation by making causal factors, their properties and the links between them visually salient, thus supporting the 'modelling-supporting-fading' strategy.

The program also mediates the uncovering of shortcomings in pupils causal reasoning, through the visual inspection of the overall casual relationships in the diagram: that is, an additive structure which combines all causes, a narrativising perspective in which causes are linked together in a single chain of events or an analytical perspective which identifies multiple interlinked causal chains.

The program's potential for supporting teachers in scaffolding learning remains untested.

Conclusions

That the evaluation 'appears to have vindicated our theory-informed design approach'.

The authors are critical of several aspects and recognise the need to have pupils explain their decision. They also recognise that the data in the log files 'has proved less promising than we had hoped' as a diagnostic tool. However, they claim that the pupils' diagrams and their presentation of these diagrams has 'demonstrated the potential of the program as a tool for diagnosing fallacies in pupils' causal reasoning'.

'In considering a more pedagogical engaged role for the computer... where the program provides coaching it must avoid misleading pupils into thinking that their reasoning is necessarily 'wrong'. Also when 'speaking for' the pupil by interpreting her actions to the teacher, the program should acknowledge the possibility that situational factors might account for apparently discrepant behaviour'.

Applying relevant theories of teaching and learning can bring order to data collected from the field and provide a cohesive framework for programme design. This should emerge through a dialogue between theory and field data which aims to ensure that (1) theories are a reflection of actual practice rather than vice versa and (2) theories embody the interplay of the social and the individual.

The concept of ZDP is insufficient if utilised alone and the modelling- supporting-fading paradigm risks ignoring the constraints of the setting.

Using this theoretical framework helped to resolve particular design issues. Reviewing field data in the light of these issues helped to identify what can be best done by the teacher and what can be best done by the pupil. It also helped to determining when and how the computer could take the role of a conversation partner.

The domain specific nature of this framework implies that software designers should develop a new framework for each domain. However, this framework could be developed to design software which fosters pupils understanding of other second order concepts.

Weight of evidence

A: Medium B: Medium C: High D: Medium

McFarlane A *et al.* (2000) Assessment and multimedia authoring a tool for externalising understanding

Setting: Primary school   **Country:** England

Broad aims

To compare what was revealed about students' understanding by a conventional test with what emerged from analysis of multimedia presentations created by the students.

Topic focus

The study looks at what can be learned from the products of pupils working with a software multimedia authoring package 'Hyperstudio' about their understanding of matters related to drug use compared with what emerges about their understanding from a conventional paper and pencil test. Pupils worked in groups to produce multimedia presentations and were tested individually after presenting their work.

Research questions

That multimedia authoring tools offer a richer view of the procedural and conceptual thinking of learners for assessment purposes.

Study design

The group of pupils who were the subject of the study were in three of the seven schools in a multimedia

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

83

project initiated by an LEA. After producing their mulitmedia presentation, they were given a test and were interviewed for qualitative findings. Therefore data from two different sources was compared in order to assess the participants' knowledge of drugs.

Data collection and analysis

Data collected by paper and pencil test and by analysing the work done in the multimedia project and from group interviews with pupils talking about their work.

Results

'Comparisons between the test and analysis of the pupils' multimedia work, and analysis of their talk about their work, suggests that the test results incorrectly gave an impression that the pupils had a thorough grasp of key facts. As a method of assessing pupil achievement the test failed to reveal a number of assumptions and associations, and preconceptions and misunderstandings, that were evident in the pupils' multimedia work and talk about that work'. 'Approximately one-third of the pupils responded correctly to all of the questions on the test and a third responded correctly to all but one question. Only one of the pupils agreed with a statement stereotyping drug users; the majority (90%) disagreed and the remaining ones indicated that they were unsure. A similar proportion agreed that 'Drugs can be dangerous if used for fun. Overall the test suggested that a substantial majority of pupils had retained knowledge about the key facts and concepts, or at least an ability to recognise them. In contrast, analysis of multimedia presentations and discussions about them suggest that the pupils had developed little understanding of a number of central concepts although they could select the correct statement about such concepts on the test'. (p 204)

Conclusions

The end products of multimedia authoring 'proved highly effective for assessing the pupils' true understanding of the complex issues involved, irrespective of the extent to which they successfully used hypertext links and pathways to connect their ideas in meaningful ways.' (p 202) 'Regardless of whether multimedia authoring software can be linked to improvements in pupil content understanding, its potential instructional value in terms of supporting pupil thinking (through formative assessment) and teacher recognition of achievement should have positive benefits for the teaching and learning process'. (p 204)

Weight of evidence

A: Low B: Medium C: High   D: Medium

Means B, Penuel B, Quellmalz E.(2000) Developing assessments for tomorrow's classrooms

Setting: Elementary and High school   **Country:** USA

Broad aims

To develop an assessment tool for problem-solving and the outcomes of students' online research.
For teachers and students to evaluate group collaboration.

Topic focus

The study is the trial and evaluation of a prototype assessment task for problem-solving. Students search the internet for information and use it to accomplish the task. The task poses the problem that a group of foreign exchange students wants to come to the US for the summer and needs to choose one of two cities. In the middle school version and the high school version, different criteria are given as a basis for decision. Students were asked to evaluate the credibility of the information on particular web pages and to formulate a search query for finding additional relevant information. Students were asked to compare and weigh all this information in making their choice and present reasons for their choice in writing to the foreign students. It takes about two hours to complete. The student outcomes assessed included technology use and reasoning with information and communication.
This paper describes the development of this instrument in the stage beyond pilot-testing but it remains work in progress.
A second development - using palm-top computers to help teacher and student observe student collaboration - is described but, as this only uses the computer as a convenient way of recording ratings, is not itself an ICT assessment tool.

Research questions

Study of the responses of middle school and high school students to a computer-based assessment of problem-solving

Study design
The middle school version of the assessment was tried out with 31 middle school students and the high school version with 62 students. Their responses on-line were rated.

Data collection and analysis
Student outcomes data were collected including: technology use, reasoning with information and communication. No details are provided of how these were collected and what form they took. Outcomes were scored by two raters.

Results
Overall, the participants handled the search and word-processing easily. The request to identify questionable information on the websites they visited was interpreted in a variety of ways. In general, students demonstrated greater proficiency at finding topically appropriate information than at reasoning with the information or communicating conclusions in a well-organised and thoughtful manner. Relationships were found between students' prior experience with technologies and their outcome scores.

Conclusions
One major advantage of embedding assessment within learning activities is the heightened focus on learning outcomes. Technology can make assessment tasks of the kinds of skills needed for the 21st century knowledge economy more feasible - providing assessment tasks that mimic the features of real-world problems and provided portable, easy-to-use templates for collecting and storing classroom assessment data.

Weight of evidence
A: Low   B: Low C: High D: Low

O'Neil H *et al.* (1999) Computer-based collaborative knowledge-mapping to measure team processes and team outcomes
Setting: Middle and High School   **Country:** Germany (but with US teachers and pupils)

Broad aims
The study examined the feasibility and validity of using a computer-based collaborative knowledge-mapping system to measure teamwork skills.

Topic focus
The study focused on the nature of the interaction between team members as they jointly constructed a knowledge map. The interest was in the nature of the interaction between team members and how that interaction influences team performances on a computer-based, collaborative knowledge-mapping task. Team processes investigated were (a) adaptability, (b) communication, (c) co-ordination, (d) decision-making, (e) interpersonal and (f) leadership.

Research questions
The hypothesis was set that computer-based, networked, collaborative knowledge-mapping systems are valid and feasible as a means of assessing teamwork skills. In addition, the expected relationship between teamwork processes and knowledge-mapping tasks were stated: (1) Adaptability: There will be no overall effect on knowledge-mapping performance. However, more effective teams will detect problems with the surface features of the knowledge-mapping task. (2) Co-ordination: More effective teams will be sensitive to the time constraints of the task. (3) Decision-making: More effective teams will use more decision-making messages than less effective teams. (4) Interpersonal: There will be no effect on performance (5) Leadership: There will be no effect on performance

Study design
Two pilot studies were conducted to assess the feasibility of the approach. The first assessed the functionality of the computer system and the second reflected some major changes to the computer system as a result of the first study.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

85

In the main study, a pre/post-test design was used to assess the reliability of the teamwork assessment. Teams of three worked to produce knowledge maps on environmental science. To collaborate, they sent messages to each other; most messages were predefined and could be sent by pressing buttons. The messages were classified as relating to five aspects of collaboration.

Performance was measured by how well a team's map compared with an expert map.

Data collection and analysis

A scoring algorithm implemented over real time was used by the computer was used to score the knowledge maps created by the teams, and to compare students' maps with expert maps. Scoring and categorisation of all messages passed between teams during the course of the construction of the knowledge map was also completed.

Results

Measuring team processes and performance using a group knowledge-mapping task was demonstrated as feasible.

The reliability of the teamwork process and teamwork performance measure was not demonstrated because the sample was too small to conduct a formal test-retest reliability analysis.

The expected relationships between teamwork process measures and teamwork performance were not found. The relationship between teamwork measures and outcomes were negative. The more messages a team sent, the poorer they performed on the knowledge-mapping task.

The teams increased their performance on all the knowledge-mapping outcome measures from the first to the second test.

All the 37 messages provided were used at some point by all teams. Decision-making was used far less than the other teamwork processes. The frequency counts of other categories were fairly uniform.

For the pre-test and overall, significant moderate correlations were observed between co-ordination and (a) decision-making and (b) interpersonal processes. For the post-test, significant moderate correlations were observed between adaptability and (a) co-ordination and (b) decision-making. For communication, which is a composite of all other processes, the pre-test data show significant and high correlations with co-ordination, decision-making and interpersonal processes. For the post-test significant and high correlations were observed between communication and adaptability, co-ordination decision-making and interpersonal processes. The pattern of correlations among teamwork processes differed by occasion with no obvious pattern of relations.

There were significant differences on pre- and post-test outcome measure between expert and novice knowledge maps, with post-tests being higher than pre-tests.

The less a team used adaptability messages, the higher their score was on the knowledge map. The number of nodes in the knowledge map in the post-test correlated negatively and significantly with adaptability.

There was no significant correlation between the co-ordination of messages and performance.

The less a team used decision-making messages, the higher their performance.

There was no significant correlation between interpersonal processes and leadership and performance.

Conclusions

A significant difference was found between low and high performing groups on the measure of reading text box measure; low performing groups focused more on reading the text box than high performing groups.

Computer trace data showed that high performing groups generated more knowledge-mapping events than low performing groups.

No significant difference was found between high and low performing groups on the use of knowledge-mapping focused messages, suggesting low performing groups did not benefit from discussions.

Students jointly using the networked knowledge-mapping system were able to jointly construct a knowledge map.

Participants had little content knowledge and participants semantic content scores were much lower than experts.

The finding of no significant correlations between most team processes and outcome measures was unexpected. This may be attributable to the low reliability of the team process scales.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

86

The task is highly knowledge dependent and participants lacked the requisite knowledge to engage with each other at a substantive level. The use of messages may reflect more on the procedural aspects of constructing a knowledge map than substantive discussion about content.

Selecting messages may have created a split attention effect; the attention demands of selecting and sending messages detracted from the task.

This raises questions about what a team is and what a teamwork task is. The collaborative knowledge-mapping task may be more like a small group task than a teamwork task. A measurement system based on teamwork processes may be insensitive to small group processes.

The computer system should be improved to lessen the attention demands of the message systems.

This is a viable means of assessing teamwork processes, given the alternatives. Existing approaches (manual) are labour and time-intensive. This software is designed to be domain independent and should transfer to other computer-based team environments, and be capable of speedy assessment of teamwork processes and outcomes in educational, industrial or military environments.

Weight of evidence

A: High B: Medium C: Medium D: Medium

Osmundson E *et al.* (1999) Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings

Setting: Elementary school   **Country:** USA

Broad aims

To investigate how computer-based knowledge-mapping could be used simultaneously as an instructional tool and an assessment tool in a classroom setting.

Topic focus

The objective of the study was to investigate how computer-based knowledge-mapping could be used simultaneously as an instructional tool and an assessment tool in a classroom setting. Knowledge-mapping was used as a repeated measure (pre-and post-) and, for the experimental group only, integrated into instruction. The subject matter was the interconnections within and between systems of the human body. The experimental group created three knowledge maps during the four weeks between pre- and post-tests. The control group worked in small groups used a computer and other materials to research the human body.

Research questions

(a) Did students' map scores improve as they studied the human body? In other words, is knowledge-mapping as an assessment tool sensitive to instruction?
(b) Did knowledge-mapping support and /or promote learning?
(c) Did use of the mapping tool foster development of understandings that were interconnected?
(d) How was the knowledge-mapper used during instruction?
'From a technical standpoint we were interested in exploring the feasibility of multiple, simultaneous online users of the mapping software as well as of providing an online automated scoring system for the student knowledge maps'.

Study design

A non-equivalent control group design was used. The experimental group worked together with knowledge-mapping software to generate collaborative online knowledge maps of the human body. The control group worked in small groups used the computer and other resources to research the human body.

Teacher was a 'veteran' instructor, with over 20 years teaching experience. She volunteered her classroom as a site for the study. She taught the two groups used in the study, one in the morning and one in the afternoon. The teacher had little computer experience. Both groups of students were familiar with a computer and with the task of knowledge-mapping.

Pre-test and post-test measures: The knowledge-mapping software, was used to assess students' conceptual understanding of the respiratory, digestive and circulatory systems.

An essay task was set as a further post-test measure in order to assess students' understanding of the respiratory, digestive and circulatory systems.

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

87

Data collection and analysis

Data collected to measure aspects as findings:

The terms and links (relevant to the circulatory, respiratory and digestive systems) used in the knowledge-mapping were identified by the class teacher and five content area experts. These terms were refined and existing work on knowledge-mapping used in addition.

Two types of mapping tasks were administered to students. An individual mapping task was administered to all students prior to the start of the instructional sequence (pre-test) and at the end of the instructional sequence (post-test). Students were given 25 minutes to complete their maps.

Using the same mapping instructions, the experimental group used the same mapping software in groups to create, review and revise their collaborative maps. They had 45 minutes for this on three occasions.

There was also a pencil and paper essay task used to assess student understanding of the three related systems of respiration, digestion and circulation.

Pre-test and post-test use of computer-mapping software which were rated on three scores:
(1) an expert content score
(2) a proposition quality rating
(3) a system interconnection score

Student essays were scored on a five-point holistic scale by six independent raters.

Results

No significant differences were found between the control and experimental groups on the pre-test map performance. All data suggested the groups were comparable.

All students improved on all measures (number of pragmatic propositions, number of scientific propositions, and number of highly principled propositions and interconnection). Overall, the experimental group showed higher mean performance on the outcome variable than the control group.

Intercorrelations between knowledge map content score and holistic essay score were of 'moderate magnitude'.

For each knowledge map measure, an ANOVA was conducted. Significant differences were found between pre-test and post-test performance for the control and experimental groups on knowledge-mapping with content score, with interconnection score, with number of pragmatic propositions, with number of scientific propositions and with number of highly principled propositions. No significant differences for the number of nonsense propositions.

Microanalysis of one group's propositions was used to investigate how students' propositions changed over time. The total number of propositions created increased from one occasion to the next 'the process of learning appeared to involve adding bits and pieces of knowledge to their system of understanding…the visual display of the monitor and its interactive nature provided a tangible visual organiser around which students could negotiate meaning'.

Conclusions

'Our data suggest that students who used the knowledge-mapping software over a period of weeks gained a deeper understanding of the relationship between the scientific concepts, both within each system of the human body and between these systems. Students in the experimental group made more scientific and also more highly principled links in their knowledge maps. In addition, knowledge-mapping helped students construct more interconnected understanding of the human body.

As well as a classroom tools, 'the mapping software clearly functioned well as an assessment device. Change in mapping scores for all students from pre-test to post-test, showed that the knowledge assessment measures were sensitive to the intervening instruction. The knowledge-mapping software, employed BOTH an instructional tool and an assessment tool, combined to make it such a useful device. By using the mapping tool in the classroom, a teacher can facilitate children's learning and so gain a better understanding of that learning by seeing how students are grasping new material and where instructional emphasis should be placed'.

'Tools such as the knowledge-mapper, with their automated scoring and ability to deliver immediate feedback to users, have the potential to serve both as instructional supports for students and teachers and as information providers for teacher about the ways in which student gradually come to refine their understandings in science.

'One way to use technology well in the classroom is to embed assessment opportunities in authentic, meaningful instructional tasks.'

Weight of evidence

A: High B: High C: High D: High

Schacter J (1997) Feasibility of a web-based assessment of problem-solving

Setting: Middle and High school   **Country:** USA

Broad aims

This was an exploratory study that assessed students learning and problem-solving in the topic area of environmental science.
To study how the process of exploration, extraction, simplification and organization contribute to problem-solving performance on concept-mapping and information-seeking tasks.

Topic focus

The authors say that the detailed reporting of students cognitive processes, while engaged in learning activities, has been a hurdle in attempts to link assessment with instruction. Even in small scale assessment, collecting and reporting the kinds of detailed qualitative analyses of students processes are tremendously time-intensive and expensive. To combat the time-intensive and monetary issues of using humans to collect and analyse process data, we had our problem-solvers work in a computer networked environment. The benefits of networked environments are that they log and store automatically all the process behaviours of those engaged at the computer. Thus, using networked technologies enabled us to overcome the hurdle of collecting, coding and storing student process and performance data which in turn enabled us to report on both how the student performed, and the process behaviours that occurred during that performance.
The study explores students' problem-solving in the domain of environmental science. The 'problem' was not stated explictly, but was presumed to be how to improve their concept maps by searching the web. Students first created concept maps on the computer, then had 40 minutes to help them improve their maps, which were then used as a post-test, the pre-post-test measure was taken to indicate the problem-solving ability in using the web.

Research questions

Hypotheses
The following hypotheses address how the processes of exploration, extraction, simplification and organization contribute to student problem-solving performance on concept-mapping and information-seeking tasks.
(1) With the addition of an electronic information space, students performance from pre-to post-test concept-mapping scores will improve.
(2) More exploratory information seeking behaviours (i.e. browsing, scan and select, and analytic information seeking) will be predictive of higher scores on finding relevant information and higher scores on final concept-mapping.
(3) Extracting more relevant information (i.e. bookmarking) from the web environment will predict higher final concept-mapping scores.
(4) High scores on metacognition will predict high scores on relevant bookmarks and final concept maps.
(5) High scores on task simplification (i.e. focusing on bookmarking concepts one knows little about) will predict higher final concept-mapping scores.
(6) Students who organize and regulate their learning by accessing feedback on their concept-mapping performance will score higher on final concept maps and finding relevant information than those who do not.

Study design

This was an exploratory study that assessed high school and middle school students' learning and problem-solving in the topic area of environmental science. The students completed concept maps on their knowledge of the field and then, after a period of using the web as a resource to explore, extract, simplify and organize the information they found, they completed a post-concept-mapping exercise. These pre- and post- concept maps were compared (in terms of scores generated by comparison with 'expert' maps) along with data from a metacognitive survey and a bookmarking measure.

Data collection and analysis

The variables below were used to measure aspects of the sample as findings of the study.
The paper details two performance measures, four process measures, a metacognitive survey and a bookmark measure. These are set as follows:
Performance measures
(1) Problem-solving performance. Our dependent measure of problem-solving was students growth from pre-to-post-test concept-mapping. The pretest concept-mapping task was designed to measure students' content knowledge of environmental science by requiring them to construct semantic relationships among

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

89

important concepts and facts. The post-test concept-mapping task was completed in conjunction with a simulated web environment that was designed to measure student problem-solving.

(2) Information-seeking performance. The dependent variable that measured information-seeking was relevant bookmarks. Relevant bookmarks told us how relevant the information students found was to helping them develop meaningful links between concepts.

Process measures

(1) Exploratory processes. All mouse clicks and pages visited during each student's search were compiled and logged electronically. The log recorded the student ID number, the machine IP address, the student map ID, the time the student spent on each web page, all the search terms the student used, each time the student accessed feedback, and all URLs the student visited.

(2) Extraction processes. Relevant bookmarks measured how relevant the information students' extracted and sent to specific concepts in their maps was. A relevant bookmark was a web document, a specific Universal Resource Locator (URL), that helped students make at least one meaningful link between two or more concepts.

(3) Simplification processes measured how well students simplified the task by attending to and focusing on which area(s) of their concept maps needed most improvement. These processes measured to what degree students sent bookmarks to concepts in their maps that needed most improvement.

(4) Organisational processes. Students' organisational processes were measured through (a) a metacognitive survey adapted from O'Neil and Abedi's (1996) Self-Assessment Questionnaire, and (b) how often students accessed the feedback. At a global level, we report frequency counts of how many times throughout the task the student accessed feedback about his or her performance. Each time a student accessed feedback, one point was awarded.

Relevant bookmark scoring: Student bookmarks (i.e. unique URLs sent to concepts) were scored against ratings of each web page in the database for each concept. The web page relevance rating rubric (Appendix B) was a five-point scale based on previous research (Baker, Aschbacher, Niemi and Sato, 1992). The 0-to-4 point scale assessed the relevance of each web document for each of the 18 concepts.

Metacognitive survey: The 32-item four-point Likert scale survey ranged in scores from 32 to 128 and assessed the constructs of planning, cognitive strategy, monitoring and searching strategies. Higher scores represented greater metacognitive ability.

Results

A pair-wise t-test confirmed the hypothesis that student performance improved with the addition of a web environment.

When a regression equation was run with total number of bookmarks as the dependent variable and browsing, scan and select, and analytic behaviours as the independent variables, these information seeking behaviours collectively predicted finding more bookmarks. Browsing behaviours significantly predicted finding more bookmarks.

Extracting more relevant information lead to better performance on final concept-mapping. Both the number of relevant bookmarks and the percentage of relevant bookmarks predicted better final concept map performance.

High metacognitive scores did not predict finding relevant bookmarks or scoring high on final concept maps. No subscales of the metacognitive survey predicted finding relevant bookmarks or scoring high on final concept maps.

A pair-wise t-test indicated that, of the sample of students that needed most work on their maps, those who focused on the concepts which needed most improvement scored significantly higher than those students who did not.

A one-way ANOVA, dividing use of feedback into three categories (i.e. never accessed feedback, accessed feedback one to four times, accessed feedback five times or more) revealed no significant results for use of feedback contributing to either higher final concept-mapping performance or a greater number of relevant bookmarks found.

Conclusions

'The results of this study indicate the feasibility of automated collection, scoring, and reporting of student complex problem-solving process data and using that data to make inferences about problem-solving performance. As we have shown, process data can be collected in inexpensive and non-time intensive ways. Further, detailed reporting of students' cognitive processes while engaged in learning activities is a feasible goal when using networked technologies.

Weight of evidence

A: Medium B: Medium C: Medium D: Medium

Schacter J *et al.* (1999) Computer based performance assessments: a solution to the narrow measurement and reporting of problem-solving

Setting Middle School   **Country:** Germany (with teachers and pupils from USA)

Broad aims

To create an authentic problem environment that to a large degree required higher order thinking processes
To measure whether or not students were engaging in their problem-solving processes and the degree to which employing these processes had an effect on final performance

Topic focus

Problem-solving was studied 'using realistic problems in realistic contexts demanding multiple cognitive processes in the domain of environmental science. The problem was to construct a concept map from a given set of terms and links. Between the pre- and post-test mapping students search a simulated web for relevant information to improve their maps.'

Research questions

That problem-solving processes would be related to outcome (specified in discussion) 'the goal was to measure whether middle students' information-based problem-solving abilities increased from the beginning of the academic school year to the end of the year and whether or not those increases could be tied back to the innovative technologies that were introduced into these schools'.

Study design

This is somewhat unclear. 68 students were tested in the fall 1996 and then had training in using certain web-based tools to search information to improve their maps. Of these, 20 were matched to 18 other students who were assessed in the following spring.
It is said that a pre-post-test design was used but scores for pre-test refer to the fall students and post-test to the spring students. It is not clear why the same students were not involved on both occasions or when the problem-soling intervention took place. It is stated that the 'fall' group were pre-tested (making concept maps) and then immediately had experiences to improve their concept maps. Which 'pre-test' scores were reported is unclear.
In the spring, a different set of students (matched to the first) were tested and their results are given as being post-test. It is not clear that these students had the same pre-test and intervention.

Data collection and analysis

Data were collected by scoring the students' concept maps, analysing the bookmarks students noted to the simulated web pages, scoring the record of browsing, searching, focused browsing and accessing feedback.

Results

'Descriptive statistics and MANOVA indicate that students significantly improved their problem-solving performance and processes from pre-test to post-test.
They also improved 'on several of the problem-solving process variables that helped them improve their performance'.
Process measures predicted outcome variables (concept map scores), apart from bookmarking.
(1) Increase in all four processes and both outcome variables from pre- to post-test.
(2) Significant pre/post difference at 0.01 level for processes of searching, focused browsing, feedback and knowledge-mapping. No significant difference reported for browsing or relevant information found.

Conclusions

'Our initial efforts to collect and report to students on performance and process data over the Internet in real time were successful. Scores increased from fall to spring as expected. Further, the process data, with one exception, predicted outcomes. 'Not only did we report to students how they performed on outcome measures, but also which thinking processes contributed to or detracted from their performance'. Thus the authors conclude that this feedback improved students' performance.'
That the intervention had the effect of increasing problem-solving processes and success at task. The authors suggest that the lack of an observed effect on relevant information in the map may be due to students' minimal use of bookmarking on which the outcome measure was based.

Weight of evidence

A: Medium   B: Medium   C: Medium   D: Medium

Tzuriel D and Shamir A (2002) The effects of mediation in computer-assisted dynamic assessment

Broad aims

To examine the effects of a computer-assisted (CA) dynamic assessment as compared to assessment by examiner alone (EO)

Topic focus

The study is concerned with the impact of the use of computers in the mediational processes of dynamic assessment, compared with the impact of a person only mediating the learning experience.
Kindergarten children were assigned to two groups, both being assessed on some seriation tasks before and after some 'mediation' which was either computer-based (described as multimedia) or 'examiner-only'. Some children used the computer for the pre- and post-tests whilst for others these were administered by the 'examiner' - the objects to be placed in order being on cards. There was a design that combined the different test forms with the mediation forms in various ways to assist interpretation of differences.
The seriational tasks were problems that required arrangement of objects on a continuum whilst controlling for one or more dimensions that are embedded within the same set of stimuli: for example, pictures of cars (different numbers, sizes and darkness of colour). The cars are given in mixed order at the bottom of the screen and the children asked to order them at the top by dragging and dropping. Icons represent the three possible task dimensions and the child should click the one they have used. If the child makes a mistake, he or she gets a graduated prompt, increasing in explicitness until full definition of the first dimension is provided.
The pre- and post-test is the Children's Seriational Thinking Modifiability (CSTM).

Research questions

Hypotheses
(1) Children who receive CA mediation will show higher pre- to post-teaching gains on the CSTM test than EO children who have mediation by only an examiner.
(2) The difference between the CA and EO groups will be more articulated in subgroups of high versus low exposure to the computer environment than the difference hypothesised for groups undifferentiated as to amount of computer environment exposure. More specifically, children who are exposed to CA throughout all phases of administration will show higher pre- to post-test teaching gains on the CSTM test than EO children who are exposed to the examiner alone conditions throughout all phases of administration
(3) Performance on the dimension of number will be higher than performance on the dimension of size and darkness.
(4) The pre- to post- teaching gains on the CSTM test will be higher in complex than in simple tasks.

Study design

The following combinations were given to groups and subgroups:
CA pre-test, CA intervention and CA post-test
CA pre-test CA intervention and EO post-test
CA pre-test EO intervention and EO post-test
EO pre-test EO intervention and EO post-test
EO pre-test EO intervention and CA post-test
EO pre-test CA intervention and EO post-test
The study design enabled two main groups to be compared: one with Pre-test, mediation and post-test all in CA and one wholly administered by an examiner (EO) (12 children in each of these groups). The effect of the mode of test administration could be examined by considering other groups who were given either pre or post-tests in each form - in all combinations (6 children in each of these sub-groups).

Data collection and analysis

Pre- and post mediation scores using the Children's Seriational thinking Modifiability test. Information was collected about number of trials and length of teaching time under CA and EO conditions. Reference also made to the complexity of the problem (although no information about problems of different complexity being used)
The data were collected through the administration of the CSTM test (Childrens Serialisational Thinking Modifiability test) as which assesses children's ordering and serialising abilities, and through dynamic assessment of their ordering and serialising capabilities.

Results
The main hypothesis – that CA is more effective than EO assessment – was examined in regard to total CSTM score, the dimensions (number, size and darkness) and complexity level. Both groups showed gains from pre- to post-test but the gains were greater for the CA group. The repeated AVOVA of groups by time indicated a significant interaction with the CA group showing higher gains than the EO group. For the total CA and EO groups, this difference was greater and highly significant. The gain for the total EO group was similar to the gain for the entire EO group for some of whom the tests were administered by computer. t-tests of the difference between length of teaching time and number of trial were not significant and so did not account for the differences; indeed the length of teaching time was slightly lower for the CA group. Differences were examined in relation to the different aspects in the seriation of the cars (number, size and darkness of colour). The improvements from pre- to post-test was higher in difficult (darkness) than in easy dimensions of the task. There were higher scores for both groups for number than for darkness and size but the difference was only significant for darkness.

Conclusions
The authors conclude that CA dynamic assessment increases pupils' performance more than EO administered dynamic assessment. This is more the case when CA 'saturates' the process; that is, all parts of the process are CA. 'While the same examiner taught both groups...the CA condition was uniquely different from the EO condition in that that the mediation principles were integrated with the unique characteristics of the DA (dynamic assessment) computer program'. Investigation of the effect of the measurement process by computer versus examiner showed no significant difference on pre-test scores. The measurement technique alone is not strong enough to have an impact on performance. Only after implementing a learning phase – in which the mediator taught the child how to solve the seriation tasks using the multimedia modality – did the CA condition have a stronger effect. When these are integrated with the unique characteristics of the computer program, the environment was more stimulating and motivating for pupils than EO. The advantages of the CA condition are qualitative not quantitative. In the CA condition, focusing and self-regulation are taken care of by the program itself, leaving the mediator with more energy to concentrate on higher types of mediational criteria. The more difficult the task, the higher the superiority of the CA condition. CA mediation is a complementary human-computer mode which combines the human qualities (warmth, relationship, etc.) with the computers systematic and controlled simulation of mediated learning principles.

Weight of evidence
A: Medium B: High C: High D: High

A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills

93