



**REVIEW**

---

**September 2005**

**The effect of formal assessment  
on secondary school Art and  
Design education: a systematic  
description of empirical studies**

*Review conducted by the Art and Design Review Group*

---

The EPPI-Centre is part of the Social Science Research Unit, Institute of Education, University of London

## AUTHORS

Professor Rachel Mason  
Dr John Steers  
Dr Dorothy Bedford  
Cynthia McCabe

## REVIEW GROUP MEMBERSHIP

|                                   |  |
|-----------------------------------|--|
| Professor Rachel Mason (Director) | Roehampton University                                    |
| Dr John Steers (Co-Director)      | National Society for Education in Art and Design         |
| Dr Dorothy Bedford                | Research Officer, Roehampton University                  |
| Dr Folkert Haanstra               | Professor, Art Education, Amsterdam School of the Arts   |
| Dr Richard Hickman                | Senior Lecturer Art Education, University of Cambridge   |
| Dr Trevor Rayment                 | Course Leader PGCE Art and Design, University of Reading |
| Dr Paul Rinne                     | Deputy Head and Head of Art, Sheen School                |

### Contact

Professor Rachel Mason (Director), Roehampton University  
Telephone: 020 8392 3009. Email: [R.Mason@roehampton.ac.uk](mailto:R.Mason@roehampton.ac.uk)

## ADVISORY GROUP MEMBERSHIP

|                                   |  |
|-----------------------------------|--|
| Dr Dennis Atkinson                | Lecturer, Art Education, University of London (Goldsmiths College) |
| Dr Dorothy Bedford                | Research Officer, Roehampton University                            |
| Sheila Dowling                    | Senior Lecturer, Art and Design, University College Northampton    |
| Dr Richard Hickman                | Senior Lecturer, Art Education, University of Cambridge            |
| Professor Rachel Mason (Director) | Roehampton University  |
| Dr Trevor Rayment                 | Course Leader, PGCE Art and Design, University of Reading          |
| Dr Paul Rinne                     | Deputy Head and Head of Art, Sheen School                          |
| Dr John Steers (Co-Director)      | National Society for Education in Art and Design                   |
| Dr Sylvia Willerton               | Education Consultant   |

### International members

|                  |  |
|------------------|--|
| Dr Doug Boughton | Professor of Art Education, Northern Illinois University |
|------------------|--|

|                     |   |
|---------------------|---|
| Dr Lee Emery        | Associate Professor, Arts Education, University of Melbourne      |
| Dr Lynn Galbraith   | Associate Professor, Art Education, University of Arizona, Tucson |
| Dr Folkert Haanstra | Professor, Art Education, Amsterdam School of the Arts            |

### **User members**

|                |  |
|----------------|--|
| Dan China      | Association of Advisors and Inspectors of Art and Design |
| Melanie Fowke  | Art and Design teacher, St Guthlac's School, Spalding    |
| Dominic Harper | Art and Design teacher, Cowes Secondary School           |
| Michele Kitto  | Art and Design teacher, The Dragon School, Oxford        |
| Nick McKemy    | Art and Design teacher, Bacons College, London           |
| Vivienne Reiss | Arts Council England                                     |
| Robert Watts   | Primary teacher, Normand Park School, Fulham             |

## **CONFLICTS OF INTEREST**

There are no conflicts of interest reported. Some members of the Review Group authored or co-authored studies included in the review.

## **ACKNOWLEDGEMENTS**

The Review Group acknowledges, and is grateful for, the support of the various academic institutions in which they are based, and of Mark Newman and Nicholas Houghton, for this work undertaken for the EPPI-Centre. They recognise that any such work must fall within the codes of practice of these institutions.

The sole source of specific funding for the review is that from the DfES, allocated through the EPPI-Centre. The Art and Design group's request for additional funding from the Head Trust was not successful.

## **LIST OF ABBREVIATIONS**

|         |  |
|---------|--|
| A-level | Advanced level GCE   |
| AS      | Advanced supplementary level examinations  |
| ARIAD   | Allison Research Index in Art and Design   |
| AT      | Attainment target  |
| AP      | Advanced Placement System – GCSE General Certificate of Secondary Education Examinations targeted at the 14–16 age group |
| BIDS    | Bath Information and Data Services   |
| BEI     | British Education Index  |
| CAEIR   | Centre for Art Education and International Research, Roehampton University   |
| CERUK   | Current Educational Research in UK   |
| DBAE    | Disciplined-based art education  |
| DfES    | Department for Education and Skills  |

|        |   |
|--------|---|
| ERIC   | Education Resources Information Center  |
| FACETS | Rasch multifaceted computer analysis programme  |
| FE     | Further education   |
| GCE    | General Certificate of Education A and AS level (examinations targeted at the 17–19 age group)  |
| GCSE   | General Certificate of Secondary Education  |
| GNVQ   | General National Vocational Qualifications targeted at the 16–19 age group  |
| HAVO   | Examinations for students in Dutch schools that prepare students for the senior secondary general education school leaving qualification and permit entry to polytechnics (where Art and Design is located) but not to university (where Art and Design courses do not feature) |
| IB     | International Baccalaureate   |
| MTEP   | Materiais e Técnicas de Expressão Plástica (Portugal)   |
| NFER   | National Foundation for Educational Research  |
| NSEAD  | National Society for Education in Art and Design  |
| Ofsted | Office for Standards in Education   |
| QCA    | Qualifications and Curriculum Authority   |
| REEL   | Research Evidence in Education Library ( <a href="http://eppi.ioe.ac.uk/reel/">http://eppi.ioe.ac.uk/reel/</a> )  |
| SCAA   | School Curriculum and Assessment Authority  |
| TGAT   | Task Group on Assessment and Testing  |
| WOS    | Web of Science  |
| WVO    | Examinations in Holland for students in senior secondary schools concerned with pre-university accreditation and preparation  |

This report should be cited as: Mason R, Steers J, Bedford D, McCabe C (2005) The impact of formal assessment on secondary school Art and Design education: a systematic description of empirical studies. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

The opinions expressed in this publication are not necessarily those of the EPPI-Centre or the funders. Responsibility for the views expressed remains solely with the authors.

© Copyright

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

# TABLE OF CONTENTS

- SUMMARY ..... 1
- Background..... 1
- Aims..... 1
- Review question ..... 1
- Methods ..... 1
- Results..... 2
- Implications..... 2
- Strengths and limitations ..... 3
- 1. BACKGROUND ..... 4
- 1.1 Aims and rationale for current review ..... 4
- 1.2 Definitional and conceptual issues ..... 4
- 1.3 Policy and practice background..... 6
- 1.4 Research background..... 7
- 1.5 Authors, funders and other users of the review ..... 8
- 1.6 Review questions..... 8
- 2. METHODS USED IN THE REVIEW ..... 10
- 2.1 User-involvement..... 10
- 2.2 Identifying and describing studies..... 10
- 3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS ..... 16
- 3.1 Studies included from searching and screening ..... 16
- 3.2 Characteristics of the included studies ..... 19
- 4. FURTHER DETAILS OF STUDIES IN THE SYSTEMATIC MAP ..... 28
- 4.1 Further details of studies ..... 28
- 5. FINDINGS AND IMPLICATIONS..... 43
- 5.1 Summary of principal findings..... 43
- 5.2 Strengths and limitations of this systematic review ..... 46
- 5.3 Implications..... 47
- 6. REFERENCES ..... 50
- 6.1 Studies included in map and synthesis..... 50
- 6.2 Other references used in the text of the report ..... 50
- Appendix 1.1: Advisory Group membership ..... 54
- Appendix 2.1: Inclusion and exclusion criteria..... 55
- Appendix 2.2: Search strategy..... 56
- Appendix 2.3: EPPI-Centre keyword sheet, including review-specific keywords .59
- Appendix 3.1: Details of studies included in the review ..... 61

# SUMMARY

## Background

There have been debates about the ways in which art education is and has been assessed both in recent times and in the past. Art is often linked with conceptions of creativity and aesthetics. Therefore, teaching and assessing this school subject area, is greatly influenced by these concepts. The influence of formal assessment of art creates polarised views: on the one hand, art – and pupils' creativity and self-expression - should not be constrained by formal assessment; on the other hand, assessment and success in national examinations can raise the profile and status of the subject. Whilst some question whether or not pupils' art should be assessed at all, it is evident that smaller numbers of them would study it in England and Wales at Key Stage 4 if this did not lead to a GCSE.

This review describes a section of the research literature on assessment in art education. The literature is summarised and described in terms of its findings; the research aims, samples, data types and collection are compared; and the findings as reported by the authors are also included. It should be noted that these have not been assessed for quality; they are described without any attempt to rate their reliability and validity.

## Aims

The aim of the review is to highlight the main theories and concepts of assessment and examinations in secondary art education.

## Review question

***What are the effects of summative assessment on Art and Design curricula, teachers and students in secondary schools?***

## Methods

The systematic review used guidelines and tools devised by the EPPI-Centre. In short, a protocol or plan for the research was drafted, including a provisional research question for the initial map of research in the field. Exclusion and inclusion criteria for the literature search were written. The protocol was peer-reviewed, revised and then published on the Research Evidence in Education website (<http://eppi.ioe.ac.uk/reel>). Research papers were searched, identified, screened for relevance and then keyworded to create an initial database. A map of research studies in the field was generated. Papers were data-extracted and assessed for quality and weight of evidence with respect to the research question.

## Results

The results of the initial round of searching elicited 2,945 titles and abstracts; 2,837 were found not to be relevant to the review as they were not topic-relevant, did not provide original data, were not written in English or were produced before 1977. Only eight studies met the inclusion/exclusion criteria of this review.

A map of this research was then created to provide useful information for dissemination to users of this review. This map describes the studies in terms of the following:

- their aims
- the research methods that they used
- the types of assessment practices which were described
- the samples used in the studies
- the data-collection methods and tools described in these studies
- the kinds of data that were collected
- a consideration of the questions and themes that arose from the eight studies

No synthesis of the studies took place.

## Implications

### Policy

A valid investigation could be undertaken to examine the impact on the curriculum of Art and Design examinations and, in particular, the extent to which external assessment dictates classroom practice. This should be undertaken by an independent body, such as the National Foundation for Educational Research (NFER), and might include a comparison between assessment and classroom practices in each of England, Scotland, Northern Ireland and Wales.

Other important foci would be to examine the following:

- the level to which art teachers' pedagogical style is test-driven or exam-driven
- how art teachers' resources influence what is taught and assessed
- how different examination options influence teaching and assessment

### Practice

Most of the research included in this systematic map was carried out by teachers and related to their practice. Although this probably caused the studies to be quite

small scale, that this was taking place is perhaps indicative of a desire by practitioners for research evidence that would guide them on the topic.

## **Research**

The research reports that were found in this review are about assessment methods, their effects on how art is taught, and the possible influences of them on art students' work. There was very little hard, empirical evidence in the studies found to support frequently conflicting claims. Although there was also a considerable amount of unsupported theory or rhetoric, there was little empirical research. This suggests that more reliable and valid research in this area of the art curriculum needs to be carried out. The Qualifications and Curriculum Authority (QCA), for example, could see it as their role to investigate systematically the impact of the various examination specifications or syllabuses.

## **Strengths and limitations**

The review identified very few empirical studies of the impact of assessment. The studies identified were quite disparate in character. This difficulty was compounded by inadequate reporting of important details of study methods and results. It was therefore not possible to synthesise the findings from individual studies to provide answers to the review question.

# 1. BACKGROUND

## 1.1 Aims and rationale for current review

The use of summative assessment is supported by a range of arguments. It is a means of raising educational standards (HMI, 1989). It communicates levels of achievement to interested parties (parents/pupils/teachers/employers, higher education/the public) and provides publicly recognised certification (Moon and Shelton Mayes, 1994). It is a means of judging the effectiveness of an educational institution (Moon and Shelton Mayes, 1994), provides educational opportunities for pupils from less privileged social backgrounds (Bourdieu and Passeron, 1990), and raises the status of a subject by giving it public credibility (Shohamy, 2001). In opposition to these claims are the arguments that governments are increasingly using test and examination results to help manage education systems, rather than to assist teachers in the process of educating students (Gipps and Stobart, 1997) and that they are reductionist (Broadfoot, 1996). Some argue that successful results are mainly the consequence of familiarisation with *assessment procedures* and that teaching is directed specifically towards answering *tests* rather than developing the skills and knowledge the curriculum intends. Moreover, there are fundamental difficulties related to the validity, reliability, impact and practicality of external examinations since these qualities may be in tension and their washback effects on classroom practice may be overlooked.

This systematic review was prompted by a concern to clarify the impact of formal assessment on curricula, teachers and students in secondary school Art and Design. From the early 1980s in England, Wales and Northern Ireland, the ongoing development of national aims, objectives and assessment criteria for General Certificate of Secondary Education (GCSE) and General Certificate of Education (GCE) examinations in Art and Design have been reported to be difficult for a number of reasons (Price, 1982). A key assumption underpinning the subject in secondary education has been that it is concerned with fostering creativity and imagination in students. Creativity is an elusive quality (Amabile, 1979; Boughton, 1994). Whether intentionally or not, Steers (1996) has argued that the 'high stakes' assessment and inspection regimes that have become increasingly common in the OECD (Skilbeck, 1992) are having a detrimental effect on Art and Design curricula because of a tendency towards increased prescription.

## 1.2 Definitional and conceptual issues

Art and Design is taken to mean the National Curriculum foundation subject so defined in English legislation for general education, examined by the recognised awarding bodies under this or associated specifications (e.g. Art and Design [Photography]) provided for students in state maintained and independent schools. The subject encompasses a range of practical activities, such as drawing, painting, printmaking, graphics, and crafts including textiles, ceramics, metalwork and information and communication technology (ICT) in which the emphasis is on developing creativity, understood as 'innovative application of

knowledge and skills' (NACCCE, 1999, p 30). It may also include art, craft and design history, and critical and contextual studies in which the emphasis is on developing (i) critical thinking and knowledge, and (ii) understanding of Art and Design theory and practice. In reviewing international literature, however, it was necessary to take new forms of practical activity and alternative nomenclatures into account: for example, 'mixed media', 'installation', 'live art', 'aesthetic education', 'visual arts education' and/or 'visual culture education'.

The terms **assessment** and **evaluation** may be used interchangeably. In this review, assessment refers to the methods used to appraise the performance of an individual student or group, rather than to make judgements about the quality of teaching and learning. **Formal assessment** is taken to mean externally set **examinations**, which are often internally assessed and externally moderated, or externally set and assessed **tests**. **Formative assessment** refers to an assessment that is part of teaching and learning, and sets out to help learning. **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. **Criterion-referencing** is an assessment system in which an award or grade is made on the basis of the quality of the performance of a pupil irrespective of the performance of other pupils; this implies that teachers and pupils be given clear descriptions (criteria) of the performances being sought. **Norm-referencing** refers to an assessment system in which pupils are placed in rank order and predetermined proportions are placed in the various grades. It implies that the grade given to a particular pupil depends upon a comparison between that pupil's performance and those of all the other individuals or groups, rather than upon the absolute quality of the performance.

**Assessment** in this context refers to those procedures required by awarding bodies, regional or national examination boards, as laid down in specifications, examination aims and criteria, mark schemes, and course guidelines for teachers, examiners and moderators. Traditional assessment instruments used to judge students' knowledge skills and understanding about Art and Design are often based on tests requiring capacities of recall and technical ability. **An assessment instrument** is understood to be **valid** when it assesses in an appropriate manner, the specific Art and Design curriculum/syllabus to which it relates (Steers, 1988). It is understood to be **reliable** when it enables students with similar abilities in different contexts to obtain the same results (Gipps and Stobart, 1997).

Performance-based tasks in Art and Design tend to be highly context-specific and do not easily allow direct testing (Beattie, 1995, p 53). A distinctive characteristic of Art and Design examinations, as opposed to standardised tests, is their extensive use of **moderation**, understood as the means by which the marks of different teachers in various centres/schools are equated with one another and through which validity and reliability is confirmed (Eça, 2002). The use of **criteria**, understood as the distinguishing properties or characteristics of phenomena by which a decision about their quality can be judged (Sadler, 1987), is problematic because they represent inexact human values. Eisner (1985 p. 2) argues that summative assessment in Art and Design more often relies on **connoisseurship**, teachers' ability to discriminate subtleties among student artworks grounded in appreciation of what they have experienced and assessed in the past. His theory of connoisseurship has been very influential in Art and Design education.

**Curriculum** is a complex concept that can mean all the experiences a child has under the aegis of the school. In this review, it refers to the content of secondary school Art and Design programmes, and the means of engaging with them (forms of art, craft and design practice and learning activities) that secondary schools offer to students (Eisner, 1979a, p 33). The review is especially concerned with the effects of external examinations and tests on secondary school curricula, students and teaching (Beattie, 1995). Since teaching and learning can be shaped by an array of social forces – including funding, political ideologies and educational consumerism (Eisner, 1979b; Skilbeck, 1985; Ball, 1995) – this was taken into account.

### 1.3 Policy and practice background

The review sought evidence of the impact of externally assessed tests and examination courses in Art and Design on curricula, teaching and students aged 11–18. Examples of such courses include the GCSE, the GCE Advanced and Advanced Supplementary Art and Design examinations and the vocational GCE in the United Kingdom, pre-university examination courses in the Netherlands, and the International Baccalaureate. Although the majority are targeted at the 16–18 age range, the review covered provision and practice for the 11–18 age range, because it enquired into the washback effects of formal assessment systems on Art and Design curricula.

Assessment in Art and Design education is widely understood to be problematic (Gardner, 1991; Rayment, 1999; Woods and O’Shannessy, 2002). Opinions as to the nature and purposes of the subject differ significantly. In the majority of educational contexts, it is characterised by a wide range of practical, critical and contextual activities. The criteria, which are intended to assist in expressing qualities considered essential in student work, contribute to reliable assessment, but are subject to various interpretations by evaluators holding different views (Barrett, 1990; Schönau, 1998, Eisner, 1999). There is widespread debate about their nature and application, and about the use of visual exemplars, weighting of criteria and moderation (Boughton, 1996).

The topic of assessment chosen as the subject for this review was identified in consultation with Review Group members. Six possible topics were discussed at a meeting of Council of the *National Society for Education in Art and Design* (NSEAD) (16 November 2002) where it was agreed to be the most pressing. At the time the review began there was teacher dissatisfaction with the English and Welsh GCSE and GCE Art and Design examinations. This was directed specifically at the perceived unreliability of the moderation procedures and recently introduced modular approach to GCE Advanced Level/Advanced Supplementary examinations (Hardy, 2002). There was concern about a negative washback effect of examination procedures on Art and Design curricula and courses for students at National Curriculum Key Stage 3 (11–14 years) (Hardy, 2002). These concerns were reflected in correspondence with NSEAD, in the education press, specialist journals and in internal awarding-body discussions (A’N’D NSEAD Newsletter, Summer 2001). At the same time, it was recognised that examinations often have very positive effects: for example, in motivating students and enhancing the place of Art and Design in the school curriculum.

Owing to active interest in this topic, an NSEAD assessment working party was set up in 2002. The work of this group informed the protocol for this review.

## 1.4 Research background

There is a significant body of research on assessment in general education (Harlen and Deakin Crick, 2002). However, the situation in Art and Design is unclear. Studies have been carried out of Art and Design examinations in the United Kingdom (Steers, 1996); Australia (Boughton, 1999), the United States (Wilson 1997), Canada (MacGregor, 1990) and the Netherlands (Schönau *et al.*, 1996), and there have been some comparative studies (Boughton *et al.*, 1996; Blaikie *et al.*, 2003). In America, where a move away from standardised testing is leading to a search for more **authentic** assessment techniques, Zimmerman (1992) identified a need for research and development on the interface between teachers and assessment procedures in particular. According to Boughton (1996), an enormous amount of research remains to be done on studio art assessment.

The EPPI-Centre Assessment and Learning Review Group carried out a systematic review of the impact of summative assessment on student motivation in 2002. The following is a summary of the findings that are of particular relevance to this review:

- Where passing tests is high stakes, teachers adopt a teaching style which emphasises transmission teaching of knowledge, thereby favouring those students who prefer to learn this way.
- High stakes assessment may create a classroom climate in which transmission teaching and highly structured activity predominate. It favours only students with certain learning dispositions and can become the rationale for all that is done in classrooms.
- High stakes tests may have a constricting effect on creativity as well as on personal and social development. They often result in teachers' own assessments being summative rather than formative.
- An education system that puts great emphasis on evaluation produces students with a strong extrinsic orientation towards grades and social status.

The first EPPI-Centre review of the impact of summative assessment on student motivation focused mainly on testing in science and mathematics. Whereas some of the above findings are pertinent to Art and Design, the visual nature of the subject and emphasis on transmission of skills and processes makes it a particular case. Assessment instruments have to allow for a wide spectrum of student achievements and cover evidence of process as well as product. Consequently alternatives to tests (such as portfolios, exhibitions and work journals/sketchbooks) are commonplace. Inappropriate design of these instruments or inconsistency in marking affects the reliability of assessment. Inadequate content fidelity, ambiguous and unshared concepts of domains, descriptors of performance, or the use of instruments covering a poor sample of artistic skills can all be sources of invalidity and negative washback.

In a subject in which creativity is highly valued, transmission teaching is probably not the norm. Moreover, the English system of GCE Art and Design examinations is both formative and summative (Steers, 1996). The fact that assessment covers process as well as product and that teacher assessors provide feedback to students on marking may go some way to alleviating the problem of high stakes assessment and performance-oriented goals demotivating students as reported in the first EPPI-Centre review by the Assessment and Learning Review Group (Harlen and Deakin Crick, 2002).

Developing a consensual view of criteria and standards is understood as one way of overcoming the difficulties of defining them in Art and Design (Steers, 1996). In the English GCE examination system, teachers are the principal assessors of student work, and their judgements are discussed with, and often moderated by, teacher examiners. According to the Assessment and Learning review, this kind of collegiality has a positive effect on curricula and students.

Discussion and research into assessment and Art and Design is ongoing nationally and internationally in specialist journals, publications and websites (QCA, 2004). Internationally, there are different national curricula and guidelines with associated assessment regimes and the expectation of this Review Group was that this would be grounded in research. The purpose of this review was to contribute to ongoing debate about Art and Design assessment policy and practice by identifying the research about impact.

## 1.5 Authors, funders and other users of the review

One of the authors of this review has been actively involved in advising the QCA through a longitudinal review of Art and Design examinations in England, Wales and Northern Ireland from 1979–2000 (QCA, 2004). Our international partners, Dr Lynn Galbraith, Dr Lee Emery, Dr Doug Boughton and Dr Folkert Haanstra were included because they are agreed experts on the topic under review.

The specific Review Group included a secondary school art teacher and an international expert on assessment and evaluation in Art and Design. The larger group was drawn from the membership of the NSEAD Council, which is nationally representative of different phases of Art and Design education. It included teachers from both state and private education, teachers with senior management responsibility, school governors, advisers, examiners, teacher trainers and a representative from Arts Council England. Education professors and lecturers from Roehampton University contributed to the review and an NSEAD web page invited contributions on the project from a wide range of users.

## 1.6 Review questions

Our primary review question is as follows:

***What are the effects of summative assessment on Art and Design curricula, teachers and students in secondary schools?***

However, as detailed later in this report, due to finding so few studies that directly addressed this question, the review primarily dealt with teachers and/or researchers' perceptions of the effects of formal assessment, or with evaluations of specific assessment models, procedures, conditions and techniques.

## 2. METHODS USED IN THE REVIEW

### 2.1 User-involvement

#### 2.1.1 Approach and rationale

The Review Group involved those who deliver Art and Design education in this review through the formation of the Advisory Review Group at the National Society of Education in Art and Design (NSEAD) (for membership, see Appendix 1.1). The group was involved in shaping the review in the following ways; identifying the review topic and refining the protocol and review question; commenting on the subject-specific keywording codes to ensure the most relevant information was extracted from the studies; assisting with data-extractions, and preparing a user perspective interface for the review.

#### 2.1.2 Methods used

The review topic was agreed at a Council Meeting of NSEAD. The members of the Review and Advisory Group responded to draft versions of the protocol developed by the smaller review team, which was circulated via email. They continued to be consulted in this way throughout the review. Use was made of expertise and information compiled by a previously existing NSEAD Assessment group, which was partly subsumed within the Review Group. A web page was created by NSEAD for the review, inviting external views. One member of the larger Review Group assisted with the process of data-extraction online. The overall progress of the review, the draft report and the selection of user respondents were discussed at meetings of Council and at an invited seminar organised by the Centre for Art Education and International Research (CAEIR) at Roehampton University.

### 2.2 Identifying and describing studies

#### 2.2.1 Defining relevant studies: inclusion and exclusion criteria

The review set out to identify as many relevant research studies as possible capable of providing data to assist in answering the review question about the 'Impact of Formal Assessment on Secondary School Art and Design Education'. To conduct the first stage of identifying potential studies, the research team discussed at length and agreed a set of explicit criteria for including and excluding studies (Appendix 2.1). The following indicates the substantive and methodological boundaries of the review, hence the kinds of studies that were included.

### **Populations**

The review was concerned with the impact of formal assessment on courses for secondary school students studying art, aged 11–18.

### **Educational phase/setting**

It was agreed that the review would only include studies covering the secondary phase of education and the impact of assessment on curriculum students and teachers within that educational setting, but would exclude independent or non-mainstream secondary education. The review did not encompass strategies for curriculum implementation, nor instructional techniques, nor teaching notes or curriculum guidelines.

### **Review focus**

The review sought research evidence of the impact of formal summative assessment on curricula, teachers and students in secondary Art and Design. It was decided not to use internal assessment and informal assessment as exclusion terms as this was likely to exclude reports containing external assessment and formal assessment.

### **Type of study**

The review focused on studies evaluating the impact of assessment on secondary school Art and Design education. Given that there has been little research in art education in general, there was concern from the outset that the number of empirical studies would be limited. However, it was agreed that studies not based upon empirical evidence would be excluded.

### **Date of publication**

Searches of sources was limited so as to identify studies conducted during the past twenty-five years, because it was generally agreed that, from this time, there was a fundamental shift in policy-making for education. There was also a concurrent increase in interest in the role of assessment in raising educational standards in both British and international literature (Stobbart and Gipps, 1997).

### **Language**

The review was limited to studies published in English.

## **2.2.2 Identification of potential studies: search strategy**

### **Search sources**

Research studies were identified from a variety of sources, using both electronic and manual methods. Some studies were identified from bibliographic databases, such as BIDS, BEI, ERIC and General Ingenta Services; Web of Science – Arts and Humanities and Social Science Citation Index; as well as journal databases, such as EBSCO (Professional Development Collection) – for example, *Arts*

*Education Policy Review*; and Ingenta Select – for example, *International Journal of Art and Design Education* and *Creativity Research Journal* (Appendix 2.2).

In addition, art specific sources such as the National Society for Education in Art and Design (NSEAD), Allison Research Index of Art and Design (ARIAD) and Art, Design, Architecture and Media Information Gateway (ADAM), as well as Educational Research specific sources such as National Foundation for Educational Research (NFER), Current Educational Research in UK (CERUK) and Education-line (a BEI service supporting educational research, policy and practice) were searched (see Appendix 2.2).

Specialist research registers, such as Index to Theses (searchable index to UK Dissertation Titles), Digital Dissertations and Dissertation Abstracts were also searched. Searching included printed books and journals, personal contacts, and reference and bibliographic lists and citations from already known reviews scanned to identify further potential citations and to cross-check the electronic searching (see Appendix 2.2). The wider Review Group, which included international members from countries known to have a particular interest in this topic, were asked to indicate studies with which they were familiar and suggest journals that might be relevant. These were handsearched. Citation searches of key authors/papers, reference lists of key authors/papers and references on key websites were used as well as personal contacts and direct requests to key informants.

### **Search criteria**

The main elements of the search criteria included combinations of terms (using AND and OR functions) to indicate that the study was about formal assessment in secondary school Art and Design education. In addition, component parts of each word were defined to refine the searching.

Decisions about particular strategies for individual sites were made after taking into consideration the facilities and sophistication of the site. Where possible, it was decided to carry out three methods of searches. The first, which applied to databases, allowed search criteria to be built up using Boolean operators. It used an algorithm based on the inclusion and exclusion terms as shown in Appendix 2.1 and Appendix 2.2.

The second method was applied to electronic databases with a facility for allowing searches to be made based on their own specific keywords. Criteria were based on keywords which were related to the three main groups of search terms (i.e. subject, assessment and level).

The third method was the only one possible for ordinary websites or electronic database sources offering limited searching. It was based on basic combinations of terms that were created by selecting at least one term from each of the three main groups of terms (i.e. subject, assessment and level). If it was allowed, then this method did include searching for phrases (e.g. National Curriculum) and truncation symbols that might have had to be adjusted according to the requirements of the site.

### **Search log**

In order to avoid duplication and identify the most successful and relevant search sites and sources, a search log was created. Details were logged to keep a record of:

- search source, URL address and search date
- search limits and search facilities
- download instructions and EndNote import filters required
- other relevant additional notes
- search criteria, number of references downloaded and name of downloaded file

An extract from an initial prototype Search Log (Appendix 2.2, table 2.2.1) shows an example of the structure of a more detailed search log used to record the searches of each site.

### **Reference management**

The Review Group, in collaboration with a member of the university learning resources centre, set up a database system using Endnote6© (a reference management software package) for keeping track of, and coding, studies found during the review. Titles and abstracts were imported electronically or entered manually into the first of these databases.

Where possible, references were exported directly to EndNote (e.g. Web of Science and EBSCO) or EndNote filters were used to automatically transfer references into EndNote (e.g. BEI and ERIC). Otherwise, hard copy printouts were screened and reference details were keyed into the database manually (e.g. Index to Theses).

### **Handsearching**

Twelve journals were searched using the search engines of the electronic databases. *Arts Education Policy Review Journal*, *Creativity Research Journal* and *The International Journal of Art and Design Education* were searched manually.

References were obtained by searching on online library catalogues, such as COPAC – University Research Libraries, British Library (listed in Appendix 2.2).

### **2.2.3 Screening studies: applying inclusion and exclusion criteria**

All studies were entered into EndNote database library 1. Once the first database was established, the entries were all screened to apply inclusion/exclusion criteria.

EndNote contains Database Field coding to keep track of the progress of a study throughout the different stages of the project. Each entry was coded in the custom fields to identify whether identified studies were to be obtained using 'get' and date, had been obtained using 'got' and date, or whether they were to be excluded and if so for what reason.

### **Stages in creating initial Reference Management Database**

1. *Obtaining references by searching online and/or downloadable sites* (e.g. Education Line, NSEAD, ARIAD, Digital Dissertations, ERIC, WOS, EBSCO, Ingenta). Most electronic databases will automatically delete duplicates found during a search, but if several separate searches were carried out within a database, then the downloaded duplicates (from within *each* site) were deleted at a later date.

2. *Entering references manually into the same EndNote database* (EPPI.enl). Duplicates were deleted from the EndNote database, first double checking that details were identical, then keeping track of additional sources from which duplicate references had been identified, and finally retaining records containing the most detail.

The database was used to keep track of the progress of each research report at different stages throughout the project. Research studies were either included or excluded at each stage of project development – inspecting titles and abstracts, obtaining complete reports for appraisal, selecting reports for keywording and descriptive mapping, and selecting reports for reviewing. The database was also used to print relevant information, such as a list of all reports to be obtained for appraisal, a list of reports assigned to a particular reviewer, or a list of all reports identified as suitable for an in-depth review.

### **Scanning reference list**

Exclusion criteria allowed for the exclusion of studies because they were not about topic (art or design), not about formal assessment, not about secondary school education, not empirical research or pre-1977. The title and abstract of each reference was scanned, and application of the inclusion and exclusion criteria identified references that were not relevant to the study and allowed those that were relevant to be obtained for further inspection (and stored in a second EndNote library file). Where there was insufficient detail, attempts were made to find additional data through the use of other search sources such as handsearching or other electronic bases.

## **2.2.4 Characterising included studies**

Keywording of screened studies was carried out on EPPI-Reviewer. The keywording strategy for educational research (2002) was used to code studies remaining after the application of the criteria, according to the country in which the study was carried out, the population studied, the age and sex of learners and type of study: additional keywords, which were specific to the context of this review, were added to those of the EPPI-Centre (see Appendix 2.3). These covered the assessment context, type of assessment components, procedures and marking, whether the study examined or raised issues about the quality of assessment, and what kind of impact was reported.

The review-specific keywords were developed through an iterative process of consultation with the NSEAD Council, Review Group members and the EPPI-Centre.

### **2.2.5 Identifying and describing studies: quality-assurance process**

#### ***Inclusion and exclusion of studies***

Three members of the Review Group independently applied the inclusion criteria to a sample of 30 studies from the first database. A comparison was made of the resulting decisions and, after discussion, agreement was reached over any conflicting results. The exercise was then repeated on a second sample of database references, downloaded from BIDS/BEI – in particular, on references containing no abstracts. Another comparison was then made of the three tables of results. Discussion identified considerable overlap. The agreed results were then scrutinised by EPPI-Centre staff.

#### ***Keywording of studies***

Pairs of Review Group members worked independently on the keywording and afterwards decisions were compared and a consensus reached. Members of the EPPI-Centre also carried out a quality-assurance role through applying inclusion/exclusion criteria to a subset of the sample of the studies. It was then decided that subject-specific keywords were necessary to enable descriptive mapping. The reports were keyworded in more depth by two reviewers each using the data-extraction tool (EPPI-Centre, 2002) independently online. EPPI-Centre staff also keyworded a sample of included studies as part of the quality-assurance process.

## 3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS

### 3.1 Studies included from searching and screening

A summary of the results of the searching and screening is given in Figure 3.1. Screening of 2,945 titles and abstracts identified 2,673 as not being relevant to the study for the following reasons:

**Table 3.1:** Exclusion criteria (N=2,673)

| Exclusion criterion                    | Number of studies excluded |
|--|----------------------------|
| 1 Not about topic – Art or Design      | 1,023                      |
| 2 Not about formal assessment          | 912                        |
| 3 Not about secondary-school education | 302                        |
| 4 Not empirical research               | 426                        |
| 5 Not in English                       | 5                          |
| 6 Pre-1977                             | 5                          |
| <b>Total</b>                           | <b>2,673</b>               |

From the outset of the searching, the Review Group were aware that the word 'art' might be used in a variety of ways as a keyword or in titles. Electronic searching was unable to discriminate and included phrases such as 'the art of'. Even with the addition of the word 'education', the search singled out papers that included the word 'art' or 'arts' in all its manifestations. This accounted for the large number of papers excluded because they were not about Art and Design. A significant number of American papers, for example, referred to language arts. Similarly, including the term 'secondary education' and its variations did not screen out papers that address art education in the other phases where none of the key terms were mentioned in the title or abstract. The word 'assessment' was also difficult to refine at this stage and gave rise to a number of papers about models of assessment, rather than research into assessment.

Our greatest problem was again one of which the Review Group became aware in the early stages of the project. Research in art education is a relatively recent phenomenon and the discussion of issues in the field in specialist journals and reports tends to be grounded in personal experience or analysis of theory. As a consequence, the Review Group experienced difficulty in identifying papers based on empirical research.

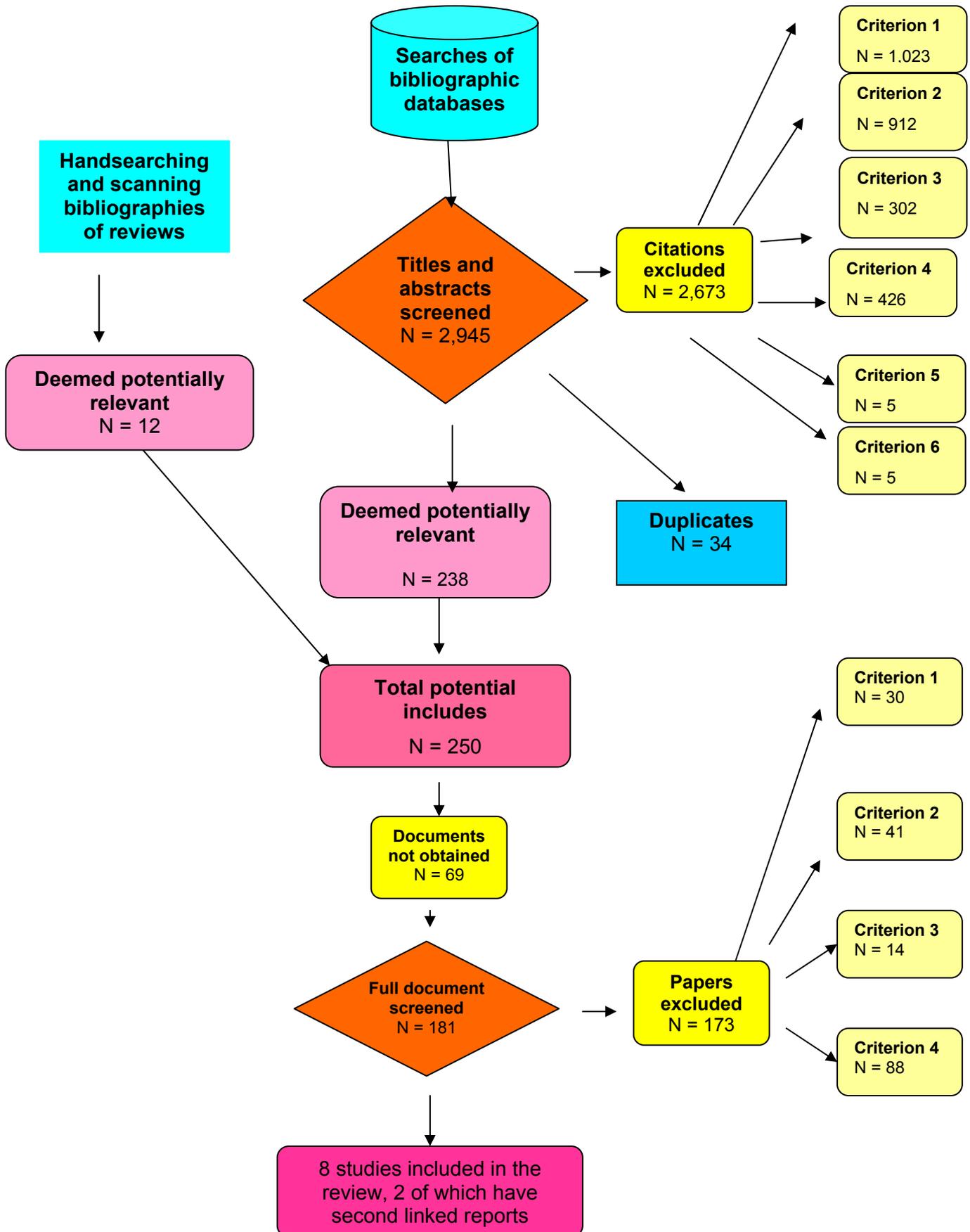
After the initial screening of papers found through electronic searching, 2,673 papers were excluded, along with 34 duplicate reports. Forty-eight

titles were found through handsearching journals and personal archives; these were screened and 12 were found to be suitable to add to the potential papers, making 250 in total.

Obtaining the full texts was delegated to a research student under the guidance of both the research assistant and members of the University Learning Resource Centre. Searching was thorough and credit should be recorded for the variety of ways employed to track down papers. A major difficulty was that many potential studies were unpublished dissertations located outside the United Kingdom, principally in Australia and the Netherlands. Electronic searching and inter-library loans enabled us to obtain copies of their full texts if they were written after 1988. The Review Group was also fortunate in having contacts in universities in those countries who tracked down and sent copies of some abstracts that appeared to have potential for inclusion. The Review Group was then able to send for full dissertations through Inter-Library loans until the cost of this became prohibitive. Our contacts sent copies of key chapters to help us to make a more informed decision. The Review Group was unable to obtain 69 documents.

Although discussion of assessment in some of these papers was linked to research, there was little or no mention of method. In some instances, references in these papers to original research reports located at universities proved impossible to track down. For example, one paper referred to a Master's dissertation at De Montfort University which university librarians could not locate. Additionally, a report referring to funded research into assessment and arts at Harvard University, which is widely discussed in the specialist field, had to be excluded after in-depth keywording, following the advice of EPPI-Centre staff that it was not a research report. The Review Group were unable to obtain the original in spite of repeated requests to the university department where the project was located.

Figure 3.1: Flow of literature through the searching and screening process



## 3.2 Characteristics of the included studies

The majority of studies were found electronically with ERIC, Ingenta and BEI being key sources. Scanning citations and reference lists was useful in that it provided a means of cross-checking the search process and identified studies which were unpublished or were academic theses. In this respect, Dissertation Abstracts, a database of dissertation abstracts, identified several studies of potential interest. The apparently unusually low number of studies found through handsearching is due to the Learning Resource Centre of our host institution having all the academic journals it holds stored electronically. Thus handsearching was replaced by electronic searching. All relevant academic journals were scanned meticulously; in particular, *The International Journal of Art and Design Education*, *Studies in Art Education*, *The Curriculum Journal*, and *Education Studies* all provided further potential studies.

The majority of the studies (five) were described in published reports appearing in journals. Two were described as unpublished PhD theses and one was an unpublished, end-of-project report. The project report and one thesis were also described in journal publications, but obtaining the unpublished reports was necessary in order to complete the in-depth keywording. One half of the studies were found in bibliographic databases and the other half through handsearching.

The paragraphs below describe in more depth the eight studies that were found relevant to this review. They are described in terms of their origins, their curricular foci, the ages of the students who were the focus of some of the studies, the types of assessments that the studies describe, the assessment components and procedures, the marking arrangements and assessment benchmarking, the impact of assessment practices and the study types.

### 3.2.1 EPPI-Centre core keywords

#### *Countries in which the studies were implemented*

Of the eight studies, six were conducted in the UK and two in North America. Two of the eight studies collected data from more than one country; one collected data in England and Portugal, and the other in England, Canada and the Netherlands.

#### *Origins of the studies*

**Table 3.2:** Origins of the eight studies (N = 8, not mutually exclusive)

|             |   |
|-------------|---|
| UK          | 6 |
| Canada      | 2 |
| Netherlands | 1 |
| USA         | 1 |
| Portugal    | 1 |

Note: Three studies reported on research from more than one country.

**Review-specific curriculum foci**

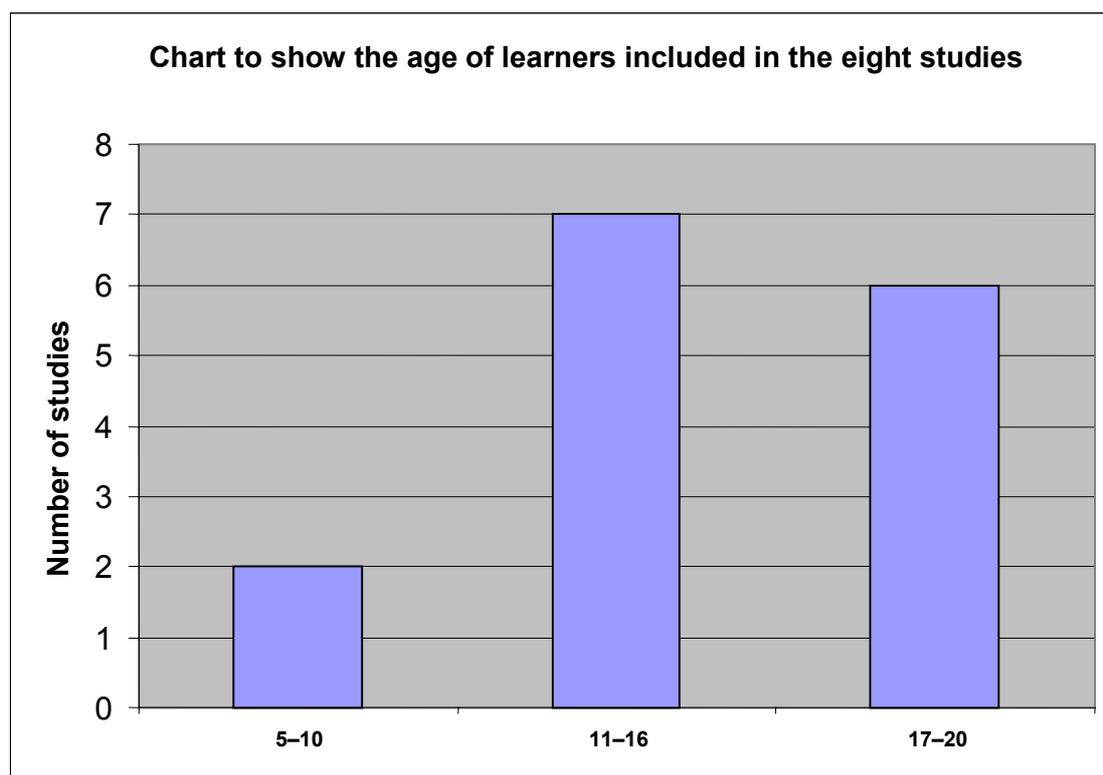
Although the focus of this review was Art and Design education, it was variously named by authors 'Art and Design', 'Art' and 'Visual Art'. One study was concerned with the 'Arts' (Art, Music and Drama).

**Table 3.3:** Table showing curriculum foci (N = 8)

|                             |          |
|-----------------------------|----------|
| Art and Design              | 4        |
| Art                         | 2        |
| Arts (Art, Music and Drama) | 1        |
| Visual Arts                 | 1        |
| <b>Total</b>                | <b>8</b> |

**Age groups of learners targeted for assessment**

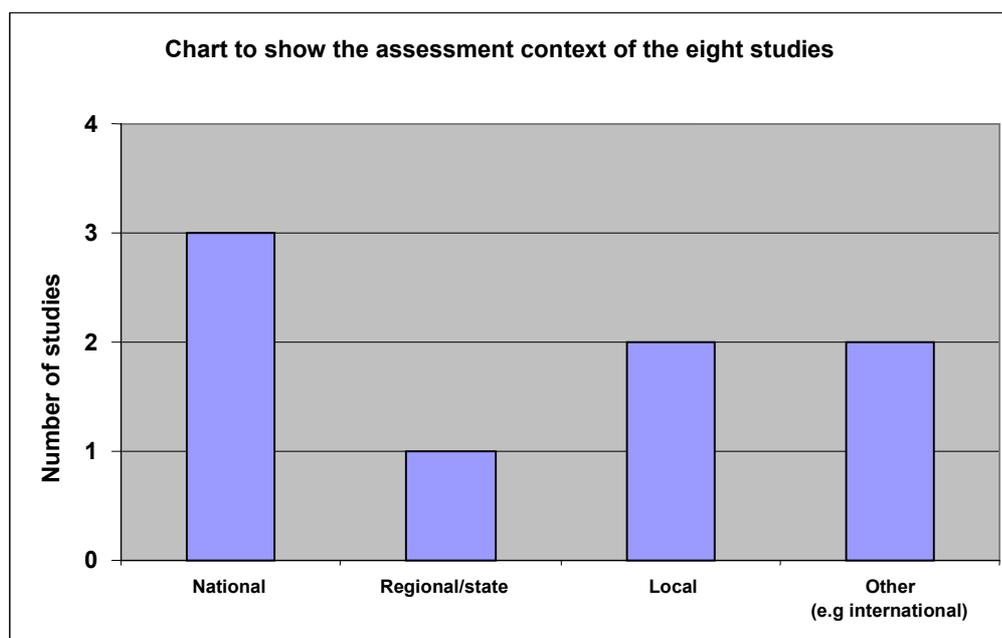
Whereas the review focuses on formal assessment of student learning in Art and Design at secondary-school level, one study focused on primary school level as well; these data were not used in the review. Two studies were of assessment targeted specifically at the 17–19 age group (GCSE A and AS levels in England, MTEP in Portugal, Advanced Placement, and International Baccalaureate and HAVO and WVO1' in the Netherlands) and were concerned with pre-university accreditation and preparation. Three studies were of assessment targeted at the 14–16 age group (GCSE) and one study was of assessment systems and procedures targeted at students in the 16–19 age group.

**Figure 3.2:** Age of learners (N = 8, not mutually exclusive)

### 3.2.2 Review-specific keywords; types of assessment

#### **Assessment contexts**

The two studies carried out in North America are national surveys of state or province-based systems of formal assessment for high school students. One study – of examinations for the General National Vocational Qualification carried out in England, Wales and Northern Ireland, targeted at students aged 16–19 – is national in scope. Three other studies carried out in England investigated local assessment practices and outcomes operating within the parameters of a national system of general educational examinations (General Certificate of Secondary Education) for students aged 16+. The assessment context for the three studies that were international in scope is national/international examination systems for students aged 18–19. Specifically, these were the International Baccalaureate (IB) and Advanced Placement Qualifications (AP) in Canada, the General Certificate of Education A/AS level examinations in England, the *Materiais e Técnicas de Expressão Plástica* (MTEP) examination in Portugal, and the HAVO and WVO examinations in Holland.

**Figure 3.3:** Assessment context of the studies (N = 8)

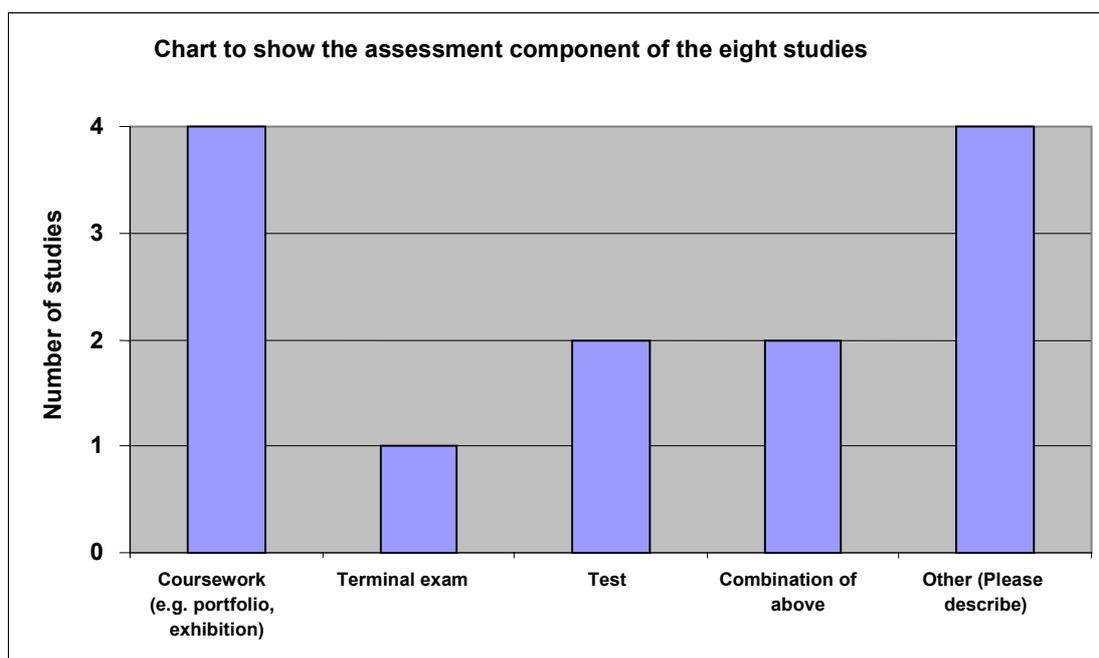
### **Assessment components**

The assessment components in the studies are varied. Coursework, in the form of portfolios or exhibitions of student artwork, features in seven studies, but in only two of these was it the sole assessment component. In the GCSE and GCE A/S level examinations in England and the HAVO and WVO examinations in Holland, that are the focus of five studies, coursework (in the form of exhibitions of student artwork) is assessed in combination with a terminal test. In the study of the General National Vocational Qualification (GNVQ) in the UK, the assessment component is coursework modules in combination with multiple-choice tests.

One study compares the relative validity, reliability and utility of an existing national examination (MTEP) dependent on an external test with an experimental examination based on an extended portfolio task which was internally assessed and externally moderated.

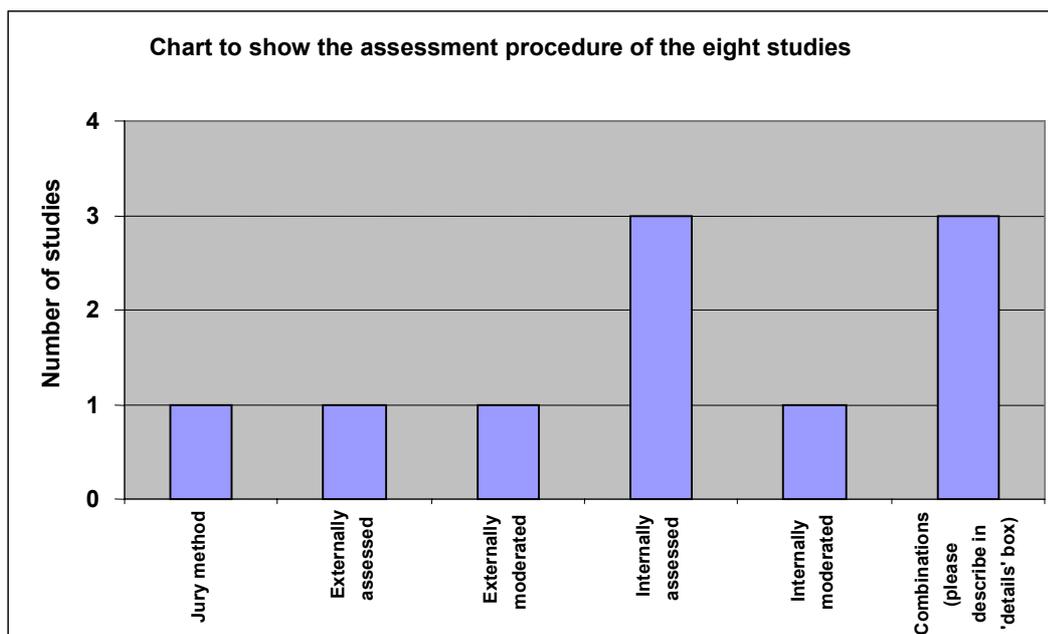
In Canada and America, there are no national art examinations. For the study located in America, the assessment component comprised visual arts tests; the study located in Canada investigated the range of components and criteria used by teachers for internal assessment. The concern was predominantly with coursework, other than in the American study, where the assessment component was stated as tests only. It was not possible to map numbers of studies dealing with either the theoretical or practical subject matter since in some they were combined. Surprisingly, given their potential to affect the validity and reliability of results, the studies did not discuss media-specific techniques and skills.

**Figure 3.4:** Assessment components of the studies (N = 8, not mutually exclusive)



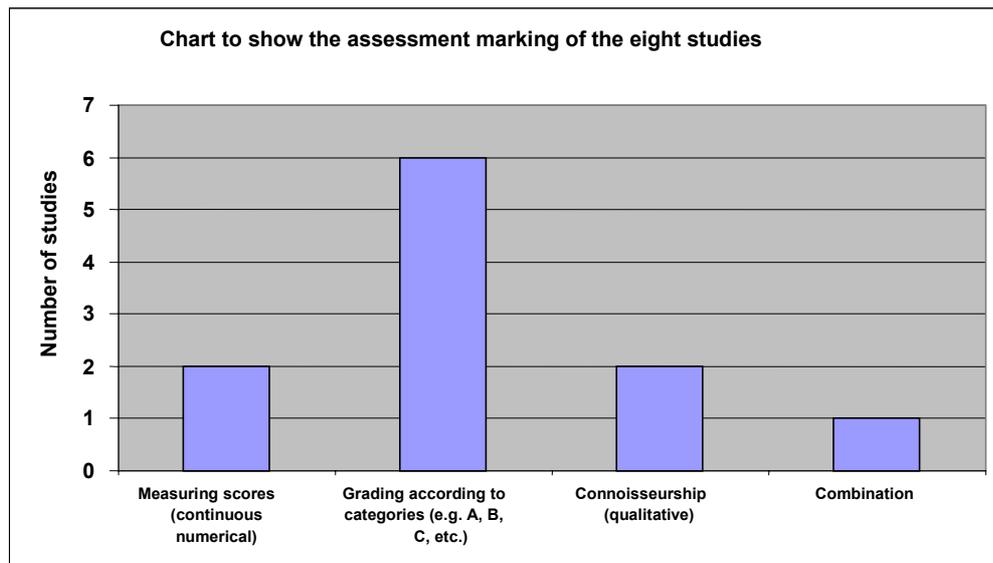
### ***Assessment procedures***

Five studies report examination systems (models) combining internal with external assessment procedures; that is, students' artwork was first assessed internally by their own teachers, after which external moderations took place. Two studies describe formal assessment arrangements, using external assessment only. In two studies, teachers' internal assessments are the study focus. In one study, the researcher experimented with a 'jury method' of assessment. On this limited evidence, it appears that assessment procedures combining internal assessment and some form of external moderation or verification have become established as a preferred model.

**Figure 3.5:** Assessment procedures of the studies (N = 8, not mutually exclusive)

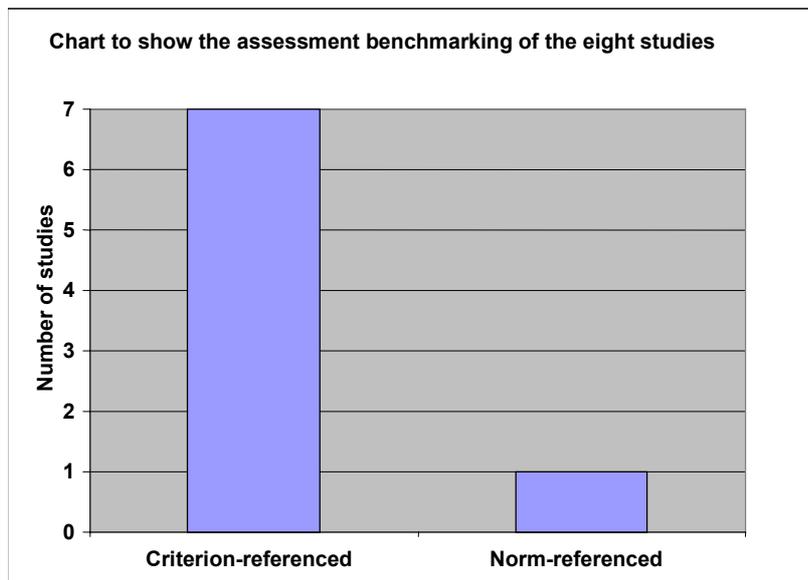
### **Marking arrangements**

Six of the studies report assessment models in which the two forms of assessment were described either as (a) coursework, or as (b) coursework and a final summative examination with student artwork graded according to categories (A, B, C, etc.). The marking system in only one study, where testing was the main form of assessment, consisted of the measurement of scores. Seven studies refer to marking systems in which it was acknowledged that judgements about the quality of student artworks were, at least partly, dependent on connoisseurship – that is, expert opinion (Eisner, 1985). Five of the studies report on assessment systems where marking contributed to the establishment of grade borderlines and assessments where students were awarded grades. In these cases, the mark schemes were linked to written assessment criteria which was an approach intended to increase the objectivity of the assessments.

**Figure 3.6:** Assessment marking of the studies (N = 8, not mutually exclusive)

### **Assessment benchmarking**

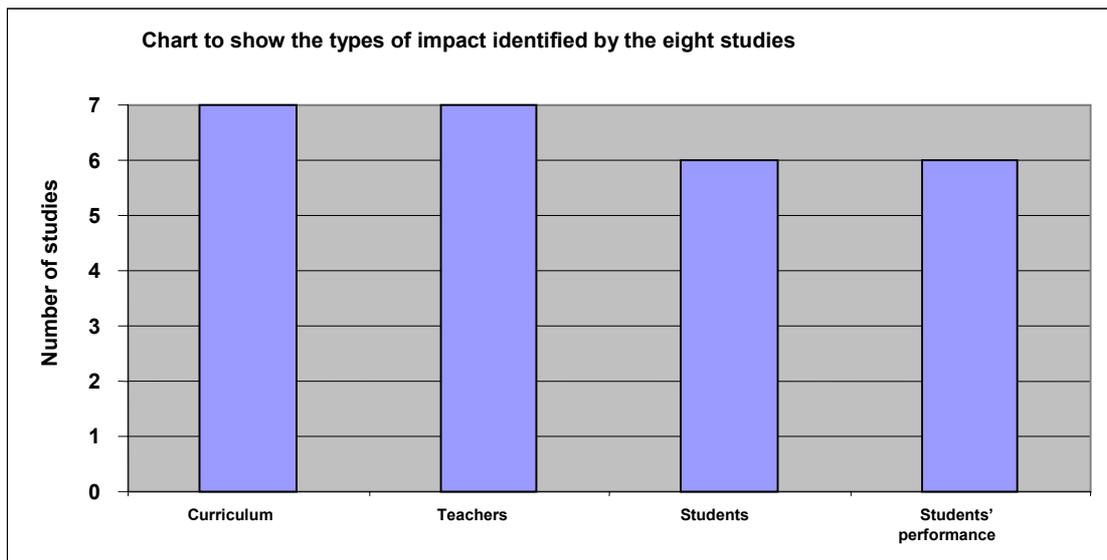
Seven studies report that criterion referencing was the accepted mode of assessment benchmarking. One study only refers to an assessment model in which benchmarking was norm-referenced. There is a suggestion in one of the studies that, although criterion referencing was official policy, it was not being applied in practice.

**Figure 3.7:** Assessment benchmarking of the studies (N = 8, not mutually exclusive)

### **Impact of assessment practices**

Seven out of the eight studies reported the impact of assessment on the Art and Design curriculum and/or teachers, and six studies refer to the impact on pupils, although impact is not a major focus in any of the studies. As a consequence, the data about impact were less specific than anticipated. The findings about impact are based on student, teacher and/or researcher perceptions.

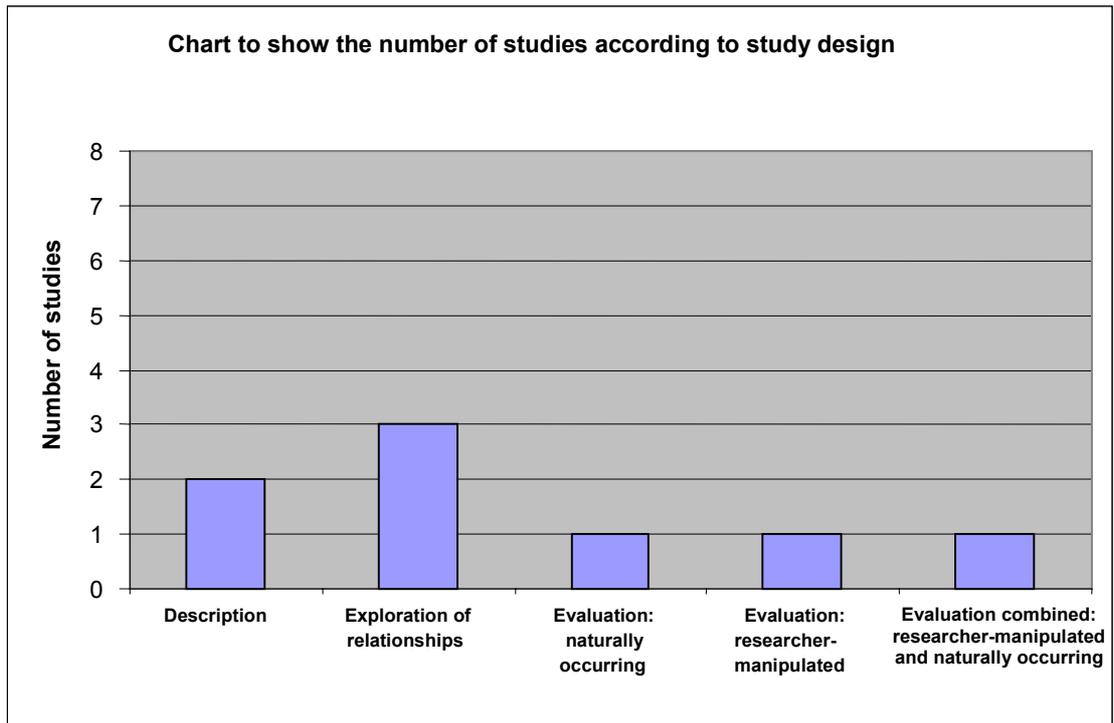
**Figure 3.8:** Type of impact identified by the studies (N = 8, not mutually exclusive)



### **Study types**

Figure 3.9 shows the three types of methodologies used in the studies. Two studies are descriptions, three are explorations of relationships and three are evaluations. In the evaluation studies, one is researcher-manipulated, one is a naturally occurring evaluation, and a third reports on both naturally occurring and researcher-manipulated evaluations.

**Figure 3.9:** Number of studies according to study design (N = 8)



## 4. FURTHER DETAILS OF STUDIES IN THE SYSTEMATIC MAP

This chapter provides summaries of the eight studies that address the review question. These include the study foci, the contexts, aims, methods, samples and populations, data collection and analysis. Findings are described as they were reported by the authors; these findings were not always substantiated by evidence or declared reliable or validated by comparison data. The reviewers recommend caution, therefore, over future citing of these findings.

### 4.1 Further details of studies

#### **Atkinson D (1998) The production of the pupil as a subject within the art curriculum**

##### ***Context***

Atkinson's (1998) study is a theoretical exposition that alluded to three case studies of the language used by teachers during oral assessment of students' drawings. Assessment discourse in the two cases relevant to this review focuses on drawings of a table set for a meal by secondary school students aged 11–12, and on drawings of chairs by students aged 15–16. The most relevant was case two, in which assessment discourse, in the form of a training video intended to assist art teachers assess GCSE coursework, was analysed. The work of Foucault and Lacan provided a theoretical basis for this study.

##### ***Aims***

Although no research questions and/or hypotheses are explicitly stated, it claimed that the principles underpinning assessment practice were in need of critical analysis. In order to show teachers' influences on students' understanding of artwork, the author sought to collect data from an analysis of drawings produced from:

- the outcome of a homework assignment (students aged 11–12)
- drawings completed in school by a small group of students aged six years
- two drawings of a chair produced by students aged 15–16 years

##### ***Study type***

The study is a description. The author observed teachers at work and examined children's work in relation to the lessons observed. The observations were recorded in the text.

### **Sample and population**

Students were drawn from Key Stage 1 (7–10 years), Key Stage 2 (11–12 years), and Key Stage 4 (15–16 years), but the study is not comparative. In one case, a boy was interviewed about his artwork. No other information about sampling is given. Reference is made to a ‘small group’ of six-year-olds and, by implication, a class of 15–16 year-olds from one UK primary school and one UK secondary school. Methods of recruitment for this study are not described. It appears to have been a small-scale study.

### **Data collection**

The author reported that he collected qualitative data from observing teachers in the classroom talking to children about their artwork. He then made judgments about what he thought were the teachers’ expectations of the students’ work as a consequence of their use of language. The author also examined the concept of drawing and questioned the ‘hegemony’ of perspectival projection. A training video was also studied. In the video, the teacher’s assessment of better and, by implication, less successful drawings, highlighted their respective formal properties.

### **Data analysis**

The analysis was qualitative, based on the author’s own judgements located within specified theory/philosophy, but there is no description of the researcher’s analytical tools or criteria for making judgements. The researcher also analysed teachers’ views in order to find any ‘subjective influences’. From these data, Atkinson speculates on the possible biases that might follow from these types of assessments.

### **Findings**

The author states that he found that assessment practices in Art and Design tend to value and produce a particular kind of drawer.

The pupil’s drawing-meaning relation may be quite different as he or she practises drawing, from the teacher’s agenda which is informed by particular traditions of signification. Reducing the hegemony of traditional forms of practice when evaluating students’ drawings, and attempting to explore how drawings might signify for students, can broaden interpretational approaches to their drawing practices and thereby help to promote a more inclusive constitution of students as drawers within their local signifying practices. (p 41)

In conclusion the author argues that, when it comes to assessing students’ drawings, the ‘...pupil is constituted within specific practices and discourses’ and that students’ perceived abilities are constructed, regulated and positioned students (p 40).

## **Bennett G (1989) Words in my eyes: the assessment of Art and Design in GCSE**

### ***Context***

This study explored, through the application of written criteria to visual material, Art and Design teachers' perceptions of the difficulties of assessing the subject. Bennett studied internal assessment of student course work for GCSE examinations. This happened shortly after the English Government had replaced norm-referenced assessment with criterion-referenced assessment, with the aim of securing provision of a fairer system. For art teachers carrying out internal assessment of student artwork on display, this meant significant changes in their assessment practices.

### ***Aims***

The aim of this study is to investigate teachers' criterion-referenced assessment practices in General Certificate of Secondary Education (GCSE).

### ***Study type***

The study is a description of Art and Design teachers' perceptions of the difficulties of assessing Art and Design through the application of written criteria to visual material.

### ***Sample and population***

The sampling methods are not described in this report; however, the study appears to be small-scale. The population focus was individual teachers and Year 11 students engaged in GCSE examinations courses. There was no comparison between the groups. The number of interviewees and observations are not specified; the study refers only to a 'small group'. The location of the schools is not specified either: they are described as students mixed comprehensive schools, one of which was in Suffolk.

### ***Data collection***

The researchers used data collected from interviews with small groups of teachers and students to investigate the move from the norm-referenced assessment of GCE O-level and CSE to the criterion-referenced GCSE. The fieldwork focused on work leading to LEAG examination in Art and Design. The findings were derived from analysis of three main types of data: (i) interview transcripts; (ii) researcher field notes of observations of teachers assessing Year 11 students' artwork; and (iii) further commentary by the participant teachers/students on the interview transcripts and field notes.

The author used a qualitative style of reporting data from interview transcripts, the examination of research field notes and teacher/student notes. The data were collected by the author.

### **Data analysis**

There was no clear description of the data analysis; it was implicit in the text. Qualitative analysis involved researcher inferences and some triangulation of the results from the different kinds of data.

### **Findings**

The author states that evidence of criterion-reference assessment carried out by single assessors lacked validity and reliability. Teachers drew on personal experience to recognise the standards they deemed necessary for allocating student coursework grades instead of the grade descriptors and stated assessment objectives supplied by the awarding body. He infers that they experienced difficulties using exemplary photographs of student artworks and the written word to make judgements, and interprets this way of assessing work as prescriptive and legitimising a particular view of art.

## **Blaikie et al. (2003) Students' gendered experiences of art portfolio assessment in Canada, the Netherlands and England**

### **Context**

The authors' rationale for conducting the study is that 'Many studies and much literature exist on the topics of gender and assessment in art education. However, there is relatively little information concerned with how high school art students experience and benefit from portfolio assessment, particularly from the point of view of gender. Acknowledging our western perspective, we have undertaken a study in which we combine findings in the literature with a small-scale survey-questionnaire of male and female students' gendered experiences of high school assessment in art in Canada, England and the Netherlands' (p 2).

### **Aims**

The broad aims of the study are to determine whether students' experiences of portfolio assessment in art are different as a result of gender and national assessment contexts.

### **Study type**

The study is 'interpretive and descriptive' (p 3). It is the result of a small-scale survey reported in relation to an extensive literature review.

### **Sample and population**

How sampling occurred is not described in the report of this study. The final sample was 107 pupils. The sample was drawn from high schools in Canada and the Netherlands, and sixth-form colleges in England, Canada (N=61), England (N=67) and the Netherlands (N=29). Sixty-four were female and 26 male (others unidentified).

### **Data collection**

The data were collected using self-completion survey questionnaires, with response data coded on a five-point Likert scale. Response categories for attitude questions ranged from 'strongly agree' to 'strongly disagree', and from 'highly accurate' to 'inaccurate' for questions about students' experience of assessment.

### **Data analysis**

Data analysis is not described in the report of this study; some results were grouped as percentages.

### **Findings**

Students in Canada, which has no national or provincial assessment systems, little moderation and fewer group critiques, had less awareness of clearly defined criteria for assessment. Dutch students were the least sure about what qualities their teachers were seeking in their work and Canadian students were most likely to view teacher-marking as unreliable. The authors cite some important features of gender differences that emerged in the questionnaire responses. The differences are as follows:

- Males were less likely than females to know and understand the qualities teachers looked for in their work: 90.6% of females and 53.9% of males.
- More female students than male students considered it important to know and understand criteria for assessment: 76.9% males thought it was important compared with 92.2% female students.
- More female students than males agreed that group critiques were useful in understanding their art work: 94% of females and 80% of males.
- Females were more open to discuss their art with teachers: 72% female and 57.7% of male students.

The authors conclude that the students' lived-experiences of assessment and national-assessment cultures influenced their preferences for certain kinds of pedagogy and assessment procedures (e.g. group criticisms and moderation). The authors state that expansion of their study would need to take place so that they would have the opportunity to look at a greater number of participants. Both the authors and reviewers commented that the evidence is based on a small and relatively uneven sample and the male reviewers suggest that the authors' interpretations are too reliant on feminist theory.

## **Eça MT (2004) Developing a new conceptual framework for pre-university art examinations in Portugal**

### **Context**

The study was carried out after dissatisfaction had been expressed by practitioners and policymakers with current Art and Design examinations in Portugal. The quality of Portuguese art examinations was evaluated in part one by means of an attitude survey targeted at art teacher/assessors and students. At

the time that the study was carried out, examinations assessed theoretical subjects (Descriptive Geometry, Theory of Design and Theory of Art and Design) and Materials in and Techniques of Visual Art. The only subject in which students produced practical work was internally assessed.

### **Aims**

- To examine the international literature about the theoretical domain of assessment in art education and evaluate its significance for Art and Design examinations in England and Portugal
- To analyse documents concerned with Art and Design examinations in Portugal and England, and identify the critical issues for assessment practice
- To observe procedures used in Art and Design examinations in England and Portugal, and identify the advantages and disadvantages of such procedures
- To investigate Portuguese students', teachers' and assessors' views of the nature and problems of Art and Design examinations in Portugal, and identify possible forms of improvement
- To compare the practices and policies of Art and Design examinations in England and Portugal, and evaluate their relative strengths and weaknesses
- To explore the implications of the study for the theoretical domain of art education and art examinations
- To suggest ways to reform Portuguese art examinations

The new assessment instrument the author developed and evaluated with Portuguese teachers in part two was a portfolio constructed over a period of one month. It could take the form of a folder, exhibition of work, work journal, CD or web pages.

### **Study type**

This was a researcher-manipulated evaluation carried out retrospectively and the researcher used before-and-after tests. Reviewers did not record details of a control group. A pilot study is described.

### **Sample and population**

The sample was taken from a population of post-16 students in Portugal and teachers in secondary schools across Portugal and England. The school students were preparing for university examinations in Art and Design subjects in Portugal. Both the teachers and students were volunteers recruited after a general announcement of the project had been made. For the pilot study, the author carried out a test of new methods of assessment in Portugal. Pre-existing differences were used to create comparison groups. Five schools in Portugal were used for the trial examination. Reviewers were unable to ascertain whether the results of this study are generalisable as the sampling was not described in enough accurate detail. Seven teachers from one school in Portugal participated in piloting this instrument with 51 students. Draft materials were discussed and negotiated in order to investigate issues of reliability and an external evaluator

observed marking trials and negotiation of marks. Then the assessment model was evaluated in a larger-scale trial in five schools, with a sample of ten teachers and 117 students.

### **Data collection**

The author first collected data from Portuguese examination documents, syllabuses and course materials, interview data with Portuguese art teachers, and observed awarding-body standardisation meetings and interviews with indicative stakeholders. The author then attended assessment meetings of English GCE A/AS level examinations in Art and Design in England, and observed assessment procedures and moderation. A procedure for external verification of internal marks, and a plan for in-service training of teachers was set up. Data collected and analysed for evaluation purposes were extremely diverse; they included teacher and moderator reports, external observers' reports, researcher observations, students' portfolios and marks, student self-assessment reports, and student and teachers' responses to interviews and questionnaires.

### **Data analysis**

Quantitative methods were used to check inter- and intra-rater reliability of marking in Portugal. In part one of the study, ten MTEP teacher-examiners marked scripts from a mock examination for Year 13 students. Nine scripts were assessed by one person on two occasions, and six of them were cross-marked by all the teacher-examiners.

The study also measured and compared the inter- and intra-reliability of MTEP marking procedures and those developed in part two. The computer program FACETS, based on the Rasch-based rating scale, was used to analyse data obtained from marking the examination scripts and trial portfolios. The authors used methodological triangulation to confirm findings from data collected by different tools.

### **Findings**

The main finding is that Portuguese teachers and students were dissatisfied with the content validity, reliability and impact of external assessment, which was perceived as limiting art content and creativity. Positive and negative evidence of the impact of the examinations on teachers, students and curricula is reported. In particular, interviewees in England stated that examination instructions, examiners' interpretations of instructions and published exemplar materials distorted and constrained coursework. They understood the accountability associated with examination league tables as having a negative impact on curriculum and as being unfair to students. On the other hand, they judged examiner and awarding-body reports to have a positive impact, because they provided feedback and professional development. The study as a whole reveals that teachers have serious concerns about reliability, validity and utility issues in both the Portuguese and English examinations, but more so in Portugal. The major outcomes of assessment, according to Portuguese teachers, were student selection, monitoring and control of the curriculum and standards, and imposition of curriculum policy. Nevertheless, for selection purposes, students in Portugal valued examination results. Teachers in both countries understood them as focusing curricula on a narrow range of Art and Design knowledge, understanding and skills.

Positive and negative effects of the new assessment model were that the model improves student motivation and independent learning, but it increases workloads. Variations in student access to resources and teacher commitment are a potential source of bias. Whilst accepting that generalisations cannot be drawn from a small-scale study, the author judged the new model of assessment worthy of further research and development in a wider context. Reviewers agreed with these conclusions, commenting that the most obvious limitations of the study were the relatively small sample in the English survey and limited trial of the new instrument in Portugal.

### **MacGregor *et al.* (1994) Assessment in the arts: a cross-Canada study**

#### **Context**

This national, Canadian survey was informed by a perceived need to improve the understanding of the criteria used to determine student achievement and to establish how they were modified by geography, experience and professional background. There were increasing demands for accountability in the classroom; university entrance requirements had started to include an assessment of arts subjects in schools and this had caused North Americans to re-examine arts programmes and the ways in which performance was assessed. There was a swing of public opinion in favour of formal, system wide assessment (p 3). The study set out to investigate concepts of teachers' learning objectives, assessment instruments, marking systems, methods of recording scores, and the use of examinations and beliefs in art education assessment strategies. The study was funded by the Social Sciences and Humanities Research Council, Canada.

#### **Aims**

The aim of the survey is to describe Canadian high-school teachers' methods of assessment in Music, Art and Drama, and whether they are modified by geography, experience and professional background.

#### **Study type**

This study is an evaluation.

#### **Sample and population**

The Canadian Education Association Handbook was used to identify school districts of a sufficient size to suggest the existence of at least one high school teaching Music, Art and Drama. From this handbook, school districts were chosen to ensure representation of school districts in every province. The authors also requested the aid of school districts across the Canadian provinces to select high schools within certain given parameters for this research and to distribute questionnaires to teachers of Music, Art and Drama. Superintendents of each school district selected were sent an appropriate number of questionnaires to deliver to schools. The logistics of questionnaire distribution meant that only schools in large urban areas were included in the sample.

The sample comprised 1,500 teachers – 500 each from the subdisciplines of Music, Art and Drama. English and French language versions of the questionnaire were made available. The target cohorts were given the option of a follow-up telephone interview. The response rate to the questionnaire was 35%, overall (527/1,500). From the telephone interviews, 10% of respondents volunteered to participate in a telephone interview; this was equally divided between Music, Art and Drama, and with francophone and anglophone representation.

### **Data collection**

The research team collected the interview data. The questionnaires were sent to headteachers/senior management to distribute to arts teachers in each subject and were returned to the researchers direct. Data were presented in tabular form (with statistics used descriptively), and as text. For the semi-structured interviews, members of the research team had a series of questions for informants which they could structure as they wished. The questions were organised around three bodies of information:

- background information about respondents (1–7)
- questions about aspects of assessment practices (8–11)
- questions related to local or provincial requirements for assessment (12–14)

### **Data analysis**

The analysis itemised and compared data on participants, their methods and means of recording assessment and their comments. Teachers' sex, language, academic background and qualifications, the level and length of teaching experience were compared against learning assignment priorities and methods, recording practices and feelings about assessment. All findings were grouped according to the subject taught and compared with each other and with practices elsewhere. Quantitative data were analysed descriptively, with tabulations presented as frequencies and percentages. Qualitative data and telephone interviews supplemented the quantitative, and were summarised and interpreted under six categories. Findings were grouped and compared both to answer the research questions and to discuss further issues for consideration.

### **Findings**

The data indicated that the assessment criteria art teachers rated most highly were developing individuality and independence, and originality of response; they also rated problem-solving abilities higher than technical skills. Cumulative assessment of individual projects carried most weight and performance-oriented examinations were prioritised over written forms. Art teachers were divided on the question of how or whether to reward effort in arts classrooms and concerned that assessment practices sometimes worked against socialisation, which they believed to be important. They preferred negotiating grades with students to student self-evaluation and voiced concern that assessment practices sometimes work against the 'struggle for empathy' in the student-teacher relationship that is crucial to the subject.

The authors state that assessment was being used to compare students (normative assessment) and did not allow measurement of individual student achievement. Moreover, pressures exerted by parents and administrators as well

as school districts affected the form and frequency of assessment. They argue that assessment in the arts is not so personal that seeking a common basis for it across Canada would be a waste of time.

### **Rayment, T (1999) Assessing National Curriculum art AT2, knowledge and understanding: a small-scale project at Key Stage 3**

#### **Context**

Rayment's study (1999) evaluated a classroom experiment in negotiated assessment and self-appraisal with 21 Key Stage 3 students. According to the author, the study followed a decision by the Schools Curriculum and Assessment Authority (SCAA) to specify two attainment targets (ATs) in the English National Curriculum, at a time when teachers were requesting more guidance on how to assess AT2 (Knowledge and Understanding of Art). Rayment states that the taught element of the intervention was constructed with the aim of establishing effective methods for pictorial communication of critical ideas. The study examined the reliability and authenticity of negotiated assessment and self-appraisal for critical and contextual knowledge, and understanding in Art and Design for secondary-school students.

#### **Aims**

This was an investigation into the evidence needed to make legitimate comments/judgements about students' critical and contextual knowledge, and understanding of Art and Design at Key Stage 3 (11–14 years). The author states that an increasing emphasis on critical and contextual work had been the most significant development of recent years in school-based art education and, as well as helping teachers to focus and organise subject content, it gave new legitimacy to the teaching of art. He therefore asked questions about (a) how students are enabled to use critical references effectively in ways that make sense and (b) about how this process is to be evaluated.

#### **Study type**

The study is a description. It describes a cross-sectional examination of a case study, using only one group of participants. Student self-appraisal and evidence from a pilot project was mentioned but not reported in detail.

#### **Sample and population**

Sampling is not described in the report of this study. The population was 21 Year 9 students who were required to produce a series of self-portraits together with evidence of process, through annotated worksheets and sketchbooks, from a specifically critical starting point. This study is not generalisable.

#### **Data collection**

Data were collected through student interview responses, student response sheets, student annotated worksheets, researcher field notes, student tasks (self-portraits, worksheets and sketchbooks), student self-assessment and teacher/researcher assessments of student work, and records of group discussion

(students and researcher). Students were required to produce a series of self-portraits and reveal evidence of process, product and critical activity for an assessment task. Common criteria, in the form of End of Key Stage 3 Statements (levels of performance), functioned as benchmarks for negotiation of marks during individual interviews with the author. The grades awarded by students and the researcher were compared with grades awarded by two art teacher colleagues and the results analysed.

### **Data analysis**

Details of the method of data analysis are scant, but there are some details about inter-rater scoring of student work involving comparison of grades (frequencies) awarded by the researcher, students and teachers. A concurrent, reflective analysis of the project was used. The outcome reported was a measure of the degree of agreement between student and teacher, and between different teachers about the grade to be awarded for a set of artwork.

### **Findings**

The study found the following:

- Students tended to underestimate their own achievements. The assessment task provided a satisfactory method for making a formal assessment, but not for assessing critical activity in routinely produced artwork which valued spontaneity and idiosyncrasy above formal evidence of critical engagement.
- Teachers expressed doubts about whether the extra time and effort (utility) involved was justifiable.
- The model of critical activity was simplistic. Activities that prescribed task, process and criteria within limited parameters ran the risk of destroying that which they set out to assess.
- Student self-assessment contributes to formative evaluation.

No evidence was forthcoming that self-appraisal improves the objectivity or accuracy of assessment. Reviewers note that teacher input and selection of the student group may have markedly affected the outcomes.

## **Sabol FR (1994) A critical examination of visual arts achievement tests from State Departments of Education in the United States**

### **Context**

The authors claim that state level testing within the United States of America had been sporadic and idiosyncratic. This study was conducted to describe, examine and critique current (at the time, 1994) state-level testing practices and to provide useful information for future creation and revision of state and local curricula and testing instruments. The study explored the range and degree of agreement of item structure and content underlying curricular themes, concepts and visual arts vocabulary, and relationships between the content and structure of the tests and related state curricula, guides or frameworks. It also examined the role of state

departments in the initiation, development and regulation of state-mandated visual arts education assessment measures. Discipline-based art education (DBAE), which identifies aesthetics, art criticism, art history and production as basic disciplines of the visual arts, provided the conceptual framework for analysing curricula content.

### **Aims**

Reviewers report the following questions in relation to state art-testing:

- What are the similarities of item and content?
- What is the range of content?
- What agreement on content themes, vocabulary and concepts from state to state?
- Is there a match between the above and those in state art curricula?
- Is there similarity of themes, concepts and art vocabulary?
- Is the curriculum in states with tests more likely to be subject-matter based?
- Do tests from states with more subject-matter based curricula have tests which reflect this content?
- Do tests from states with child-centred, society-centred or subject-centred art curricula reflect this orientation?

### **Study type**

This is a cross-sectional survey that describes and compares state art achievement tests.

### **Sample and population**

A representative sample was drawn from personnel in State Education Departments in the United States of America. Although the author does not claim the sample is representative, the reviewer considers that it is, because having contacted 50 state departments, the response level where 34 states responded and provided in excess of 100 test papers for consideration, is considered to be representative. There is no indication of the location of participating states. There were no comparison groups.

### **Data collection**

Data were collected from the following:

- art assessment tests developed by each state
- art curricula documentation (frameworks and guides) from each state
- questionnaire scripts completed by state departments
- telephone interview scripts (between researcher and state departments)

The researcher collected the data, which was stored at Indiana University.

### **Data analysis**

Data analysis comprised quantitative content analysis methods as recommended by Borg and Gall (1989). Content analysis was selected because it is used generally to produce descriptive information, cross-validate research findings or test hypotheses. Absolute and relative frequencies were used where appropriate and terms such as mean, median, mode and percentages summarised content and central tendency.

### **Findings**

The main finding is that there was a great deal of diversity in test structure, content and items, although a degree of agreement about content. There was evidence that visual art tests do not adequately address the array of knowledge, skills and processes found in visual art curricula. For example, they do not assess aesthetic content and knowledge of non-Western art. Whereas tests using a variety of item types (such as multiple-choice, essay completion and short answers) provide a more comprehensive indication of student knowledge than those using a limited selection, the author found that there is evidence that these tests do not tap into higher-level thinking or problem-solving skills; they provide only partial indications of student learning. In consequence, the study found that the limitations of achievement testing were causing some states to consider more authentic assessment measures (portfolios, critiques, reflective journals), along with achievement tests, in the hope of creating programmes that can measure student learning more comprehensively. The author added that the objective tests and production activities, which made up nearly half the components, provided only partial indications of student learning. He also found that pressures exerted by parents and administrators, as well as school districts, affected the form and frequency of assessment.

## **Willerton S (2001) GNVQ art and design: the development, distortion and subsequent dismantling of its more radical features**

### **Context**

Willerton carried out a longitudinal study of the assessment procedures for General National Vocational Qualifications (GNVQ) Art and Design. She participated in the assessment system under investigation as an examination co-ordinator/teacher, local education authority advisor, chief examiner, and awarding-body consultant, evaluating the impact of the trials of alternative assessment models on staff and students in her own school over a period of six years.

According to the author, GNVQs consist of series of vocational work units divided into elements: the performance criteria. Awarding bodies identify possible sets of units, but there are mandatory core skills units in Communication, Application of Number and Information Technology that are common to them all. The qualifications apply a competence-based model of assessment (in which documentation of evidence is intrinsic) and performance criteria, range and evidence indicators necessary to achieve an award are specified in advance. The skills, knowledge and understanding that students must achieve are set out in

detail in Vocational and Key Skills units. The primary evidence for assessment is based on a student portfolio made up of projects and assignments compiled throughout the course. Internal assessment was externally verified and externally set, and marked tests administered on three occasions in the year. A final summative assessment of unit scores results in the award of Pass, Merit and Distinction grades, indicating levels of achievement.

### **Aims**

The aims of the study are to identify and evaluate the extent to which GNVQ fulfilled the initial goals of the Art and Design vocational area, and how far increasing demands for external assessment has affected the original intentions over the period 1992–98.

### **Study type**

This is a descriptive longitudinal study.

### **Sample and population**

The researcher identified a group of students who were part of the GNVQ Centre (Centre A) where she worked as an administrator and teacher. She carried out her research at the Art and Design centre where she worked. Thirty secondary-school students were studied in order to examine why they chose the GNVQ, what they thought about the course content, and how they fared through GNVQ. Student and group achievements in the 21 other centres were also compared statistically, using national data about GCE A-level student results.

### **Data collection**

The primary data collection involved qualitative measures. The data were collected from 30 case studies of students, observation of teaching practices, interviews with students, questionnaires, and comparisons with data from students in other Art and Design centres. The researcher collected data about the impact of the trials of alternative assessment models on staff and students in her own school over a period of six years. Observation focused on how teaching content and goals were being modified in response to external assessment requirements, and how students organised preparatory work and managed time for external tests.

### **Data analysis**

There are no details of data analysis in the report of this study.

### **Findings**

Results showed that disparities were identified between GNVQ courses in the 21 centres and variations in allocation of time and resources. Moreover, the external verifiers' interpretations of procedures and demands were inconsistent. The study found little evidence of progression through GNVQ from Foundation GNVQ (Level 1), and some evidence of successful progress from Intermediate GNVQ (Level 2) to Advanced GNVQ (Level 3) qualifications.

The reviewers state that the author found the GNVQ examination system was not as successful as GCE A-levels as a means of helping students gain entrance into higher education. There was an emphasis on business skills rather than creativity of students. Teachers did report that students were more confident about their work and portfolios under this system. However, for teachers, the emphasis on external assessment seemed overwhelming and time-consuming, and so took away from traditional teaching time. The many tests were viewed as obstacles rather than as avenues for learning. In the course of the study, the style of the end-of-unit tests changed, from short-answer questions and examination papers to tests based on multiple-choice questions, but these still had an adverse impact on course delivery. The teachers and author expressed concern that the generic nature of the grading criteria militate against creativity. Although units can be assessed separately, to allow for credit accumulation and transfer, the grading criteria only applied to the full award.

## 5. FINDINGS AND IMPLICATIONS

### 5.1 Summary of principal findings

#### 5.1.1 Identification of studies

The review set out to identify research studies concerned with the impact of formal assessment processes on secondary-school Art and Design education. A broad search was made which resulted in 2,945 references from bibliographic databases, and a further 52, considered to be of potential relevance, from handsearching, citations or bibliographies.

Some study titles lacked focus and the quality of abstracts was poor. This made the identification of potential studies from the initial database difficult. However, after screening, 243 studies met the inclusion criteria; this included studies described as unpublished reports. This number of studies represents about 8%, which seems a small proportion of all citations screened. This was in part because the number of citations obtained was misleading. 1,023 citations (35%) were excluded because they were not focused on the subject area of Art and Design education. Indeed, accurate location of potential studies electronically was made complex because the search terms we used are notoriously difficult to define. For example, the term 'art' is used in the research as a proper noun denoting a subject in the school curriculum. Databases are unable to make this distinction and in many of the studies located it was most often as part of the adjectival clause 'the art of'. Redefining the term with the addition of the word 'education' still allowed for the inclusion of phrases such as 'the art of mathematics education'. The search term 'Art and Design' was unhelpful because it is specific to the British education system and identifies large numbers of studies focusing exclusively on design. Nine hundred and two were excluded because, although they were on topic, they were not about secondary-school education. Three hundred and ninety two were excluded because, although they were on topic and addressed the secondary phase of education, they were not research. Many of them were policy documents specifying curriculum and assessment standards, or curriculum exemplars offering instructional guidelines for teaching Art and Design. In hindsight, the exclusion terms could have been used more effectively in searching to reduce the number of studies which were about instruction or policy.

Although we were looking for research studies about the impact of assessment on curriculum and pedagogy, we did not feel confident in narrowing our search to this level of specificity. This lack of confidence was also based in our experience of searching for studies in Art and Design education prior to this review. As Harden *et al.* (1999) suggest, research databases in the social sciences (and especially in Art and Design education) do not yet consistently code studies according to research design. We were unable to use a study filter on bibliographic databases, despite the fact that we were looking specifically for studies about the impact of assessment in secondary education that were research-based. In hindsight, it might have been advisable to include studies reporting on primary education, but, at the time the decision was made, we had not appreciated how few studies we might find.

Searching systematically meant covering as many sources as possible and, perhaps more importantly, recording the search process. The state of the art of entering educational research into databases lags behind that of other disciplines and this is especially the case in Art and Design. Thus it was necessary not only to search all relevant electronic databases, but also to follow up citations in earlier reviews and in obtained papers, and to handsearch journals held in the library, as well as those online, which were very limited at the time of this review. Of course, no search can be comprehensive, but strength of this methodology was in the recording of details so that its limitations were made explicit: for example, dates of journals that were handsearched and procedures for searching databases. Having a team of 'experts' in the field and 'user members' involved in the various stages enabled a more thorough search, as they were able to give personal recommendations and also to look for major omissions.

We were fortunate to be assisted by a consultant with extensive experience in training researchers in the use of databases for searching and implementing techniques for reference management. Under her tutelage, search terms were applied consistently and systematically, data sources were logged and progress in obtaining studies was recorded. After crosschecking the first database, by taking a specified list of studies obtained from citations and personal references, we found that the only ones not included were unpublished conference papers or articles in press. This highlighted for us the importance of using citations, but also gave us the confidence to assume that a thorough search had been achieved. Quality control effected by team members and EPPI-Centre staff was particularly helpful, and enabled clarification of thought processes and instances where 'insider knowledge' was being used – for instance, in the process of screening the first reference library.

Work on exhaustive searching, within the context of systematic reviews of educational research, is in its infancy. Although there are developments and improvements to sites on an almost daily basis, we found inconsistencies in ways in which search terms are applied, the range of mechanisms for downloading data, reference management and the quality of information about each entry. There is little empirical research that evaluates the utility of methods of, and sources for, searching for educational research (e.g. which and how many bibliographic databases need to be searched?). However, we were able to draw on principles developed for systematic reviews in other subject areas to maximise the sensitivity of our search. Searching multiple sources was crucial; a reliance on a single source, or one or two bibliographic databases, could have missed a large proportion of existing studies. We found that searching databases had to continue until such time as no further studies were being obtained.

Although it is not unusual in systematic reviews to scan through large numbers of citations to identify a much smaller number of relevant studies, the processes involved in scanning large numbers are time-consuming. This underscores the need for more work on searching for educational research. Many university research libraries are in their infancy, particularly in the United Kingdom, but they are developing and much more is becoming available electronically. However, application to individual universities was fruitful, especially in America.

### 5.1.2 Mapping of all included studies

Within the limits of our search strategy, our map uncovered the nature of a potential evidence base about the effects of formal assessment practices on Art and Design curricula, teachers and students in secondary schools generated between 1977 and 2003.

The majority of the eight studies were conducted in England (six) with two in North America; two studies are comparative and were carried out in two other European countries also (Portugal and the Netherlands). This differs somewhat from the results of an EPPI-Centre review of the impact of summative assessment and tests on student motivation for learning in general by Harlen and Deakin Crick (2002). They found 27 studies in America, 10 in the United Kingdom and none in other European countries. However, our figures may still reflect a bias in the sources searched towards studies published in the United Kingdom and United States and in the English language (only five studies were excluded on the basis of language.)

The research team was conscious of the importance of attending to national contexts in mapping the evidence base on effects of assessment. There were interesting differences in assessment procedures, components and marking in Europe and North America that might be partly explained by different organisations of assessment systems and theoretical models. External testing in Art and Design, involving measurement of scores appears to have been discredited in many countries other than America. With one notable exception, the majority of studies refer to assessment of student coursework in the form of exhibitions and/or portfolios in combination with terminal examinations. In two studies, portfolios are a specific research focus. Commonalities across the studies are that, in the majority, the assessment procedure is internal assessment of coursework by teachers, student work is graded according to categories, and assessment benchmarking is criterion-referenced. There are fewer references to impact on students than on teachers and curricula, and the evidence lacks specificity, or is slight because it is not the main or only study focus.

Evaluation studies are most likely to produce the evidence needed to answer this review question. However, two studies are descriptive and three explore relationships. Of the three evaluation studies, it is important to note that only one involves a control or comparison group which can help to rule out the possibility that effects of assessment are simply due to unknown differences between schools, learners or teachers. Six studies use the combined strengths of collecting both qualitative and quantitative data. Two of these collected qualitative data to examine process alongside quantitative estimates of impact. The majority of studies were carried out for research degrees in higher education and the small sample sizes may reflect this and/or the lack of funding available for research in the subject area.

Our map suggested a tiny potential evidence base for specifically informing the field about the effects of assessment on curriculum and pedagogy in Art and Design. (Only three of the eight studies were judged by reviewers to be directly on topic.) There is clearly a gap for research commissioners in the Art and Assessment sector.

## 5.2 Strengths and limitations of this systematic review

### *Search strategy*

The screening of studies in the first database identified a significant number of studies that appeared to have potential to provide evidence, so the low numbers meeting the inclusion criteria was a concern. Possible explanations are that there really are very few studies of the impact of assessment on secondary-school Art and Design, or limitations in the search. Since the search strategy was scrupulously quality-controlled by EPPI-Centre staff and found to be sound and consultation with Review Group members revealed that the low numbers were no surprise, the first explanation appears credible. It is probably always going to be the case that reviews of Art and Design education yield relatively low responses. However, research into wider aspects of curriculum and pedagogy probably exists that would have yielded more discussion of impact.

### *Problems retrieving studies*

The team remains concerned about the procedures used to access studies after screening, since a total of 60 studies were not retrieved. There may be more effective means of finding studies. Due to inexperience, the amount of time and effort needed to retrieve them all was underestimated. When experts in the field were requested to send relevant studies, some of them failed to respond. Members of the Review Group identified two important studies for which the original research reports could not be tracked down (Lindstrom 1998; Beattie, 1994). There may be more effective ways of finding studies, which are not so accessible, and one relevant study was identified after the data-extraction stage in the review. These are weaknesses of the review. Other limitations may be (i) the definition and scope of the review; (ii) errors of keywording and data extraction; and (iii) the ability of the review findings to inform policy and practice. These are discussed below.

### *Problems of definition, timing and scope*

Defining the research question was difficult and it went through three formulations. The users who established the focus of the review may have had a hidden hypothesis that assessment systems have a negative impact on curricula and pedagogy. Moreover, the peer referee doubted that it was possible to separate effects of assessment from other kinds of effects and warned of the lack of quasi-experimental studies with control groups. One respondent suggested the question was dated, given that research has already shown that teaching for tests inhibits development of deep learning. In hindsight, it might have been better to phrase the question another way.

The small number of studies and inexperience at formulating review-specific keywords limited the scope of the mapping exercise. Whereas a number of interesting differences between studies carried out in North America and the United Kingdom were found, they did not assist with answering the review question. As far as the studies focusing on Art and Design examinations in England, Northern Ireland and Wales are concerned, the lack of any studies dealing directly with impact and low weightings they were afforded in terms of

research method, restrict the observations it is possible to make about the development of policy and practice.

Completing the data-extraction of the studies took much longer than anticipated and the quality varied. Since the process was unfamiliar to Review Group members and time-consuming (each study taking a minimum of seven hours), some of them lost interest or omitted to answer significant questions. EPPI-Centre staff pointed out errors and inconsistencies but it is possible these were not completely ironed out.

### ***Ability to inform policy and practice***

A final limitation of the review is related to the fact that it was difficult to obtain feedback from group members other than academics and teachers. The review might have been different had it been driven by the concerns of parents and/or students rather than teacher/assessors.

## **5.3 Implications**

### **5.3.1 Policy**

It seems extraordinary that so little empirical research exists on the impact of Art and Design examinations on classroom practice, given the concerns that have been expressed for many years. If the awarding bodies or regulatory bodies, such as the Qualifications and Curriculum Authority (QCA), have undertaken such systematic research (and there is no evidence that they have), it is not available in the public domain.

An investigation should be undertaken into the impact of Art and Design examinations on the curriculum and in particular into the extent to which external assessment dictates classroom practice. This should be undertaken by an independent body, such as the National Foundation for Educational Research (NFER), and might include a comparison between assessment and classroom practices in each of England, Scotland, Northern Ireland and Wales.

### **5.3.2 Research**

The findings of the review have implications for primary research and for future systematic reviews. Perhaps the most significant finding is that, despite an extensive literature that discusses the perceived problems of assessment in Art and Design, there is very little hard, empirical evidence to support frequently conflicting claims. There is a considerable amount of unsupported theory or rhetoric as this review revealed, but there is little research employing systematic observation and experiment. In particular, it seems surprising that neither the awarding bodies, nor apparently QCA, see it as their job to investigate systematically the impact of the various examination specifications or syllabuses.

### **Primary research**

The Review Group were aware of a lack of research in Art and Design education from the start of the exercise and anticipated searching general education studies also. However, they were advised by the EPPI-Centre to limit the review to studies in the specialist field. During the screening process, EPPI-Centre staff challenged the team's interpretation of research and the largest number of studies was excluded because they were not empirical research. Although the small number of studies available for in-depth review was a surprise to EPPI-Centre staff, it is consistent with the team's perception of the field as preoccupied with visual rather than verbal modes of communication and expression, and favouring theorising over research. The absence of research studies by well-known experts on assessment in Art and Design was a surprise. Discussion of assessment in the field clearly needs to become research-based rather than speculative.

Since the three studies providing the most evidence were all relatively recent and a study of the impact of standards referenced assessment in visual arts in New South Wales was published after the review (Andrews, 2003), it may be a growth area. But this review identified a clear need for research into the impact of assessment on curriculum and pedagogy.

A finding of the review was that the quality of research and research reporting in Art and Design is a matter of concern. The majority of studies were descriptive and there is a case for more evaluation studies using quantitative methods and studies that explore relationships. It was difficult to screen, keyword and data-extract studies because basic information about the research context, sampling and methods of data collection and analysis was lacking. Research needs to be reported more coherently, systematically and thoroughly.

There are implications from the review also for higher education providers of research degrees. Masters dissertations are seldom available. Although doctoral theses were the best source of potential studies, they were difficult to track down (with the exception of theses from the USA). Screening abstracts was time-consuming because they were poorly constructed and omitted to report results. This suggests that quality control of supervision of research studies and mechanisms for recording results need to improve.

### **Systematic reviews of research**

Reflections on the strengths and weaknesses of this systematic review suggests the following implications for future reviews in this or other areas of education:

#### *Managing reviews*

- Choosing membership of the Review Group is very important, especially in arts and other education fields where the critical mass of researchers is low. Enthusiasm for research and willingness to learn are crucial.
- For a number of reasons, including changes of key personnel, this review got very behind time. Lessons learned are that research officers need good IT skills and should systematically record everything they do. Director and co-director involvement is crucial in ensuring participation of all team members and dissemination of results.

- Practical problems this review experienced could have been alleviated with additional funds. However, the team's attempts to obtain support from arts and arts education foundations were unsuccessful, perhaps because they are not used to funding research.

***Reviews of assessment in Art and Design***

- There may have been in-built bias in the review question. There is no perfect solution and all assessment systems are fallible. The argument is about specifying good criteria in as precise a way as possible. Future systematic reviews that look for evidence of how assessment promotes learning might be more profitable.
- The scope of the review question could be extended beyond secondary education. Studies in primary and higher education with potential to provide evidence were identified in this review but had to be excluded.

## 6. REFERENCES

### 6.1 Studies included in map and synthesis

Atkinson D (1998) The production of the pupil as a subject within the art curriculum. *Journal of Curriculum Studies* **30**: 27–42.

Bennett G (1989) Words in my eyes: the assessment of art and design in GCSE. *Educational Studies* **15**: 241–249.

Blaikie F, Schonau D, Steers J (2003) Students' gendered experiences of art portfolio assessment in Canada, The Netherlands and England. *Studies in Art Education* **44**: 335–349.

Eça MT (2004) Developing a new conceptual framework for pre-university art examinations in Portugal. Unpublished PhD thesis. Surrey: Roehampton University.

MacGregor RN, Lemerise S, Potts M, Roberts BA (1994). *Assessment in the Arts: A Cross-Canada Study*. Vancouver: Visual and Performing Arts, University of British Columbia.

Rayment T (1999) Assessing National Curriculum art AT2, knowledge and understanding: a small-scale project at Key Stage 3. *Journal of Art and Design Education* **18**: 189–194.

Sabol FR (1994) A critical examination of visual arts achievement tests from State Departments of Education in the United States. Unpublished PhD thesis. Bloomington: Indiana University

Willerton S (2001) GNVQ art and design: the development, distortion and subsequent dismantling of its more radical features. Unpublished PhD thesis. Brighton: The University of Sussex.

### 6.2 Other references used in the text of the report

Amabile TM (1979) Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology* **37**: 221–233.

Andrews S (2003) The formative evaluation of standards-referenced assessment in visual arts. *Australian Art Education* **26**: 34–41

Ball SJ (1995) *Politics and Policy Making in Education: Explorations in Policy Sociology* (3rd edition). London: Routledge.

Barrett M (1990) Guidelines for evaluation and assessment in art and design education 5 to 18 years. *Journal of Art and Design Education* **9**: 299–314.

Beattie DK (1994) The mini-portfolio: locus of a successful performance examination. *Studies in Art Education* **47**: 14–18.

- Beattie DK (1997) *Assessment In Art Education*. Worcester, Massachusetts: Davis Publications.
- Blaikie F, Schonau D, Steers J (2003) Students' gendered experiences of art portfolio assessment in Canada, The Netherlands and England. *Studies in Art Education* **44**: 335–349.
- Borg W, Gall M (1989) The methods and tools of observational research. In: Borg W, Gall M (eds) *Educational Research: An Introduction (5th Edition)*. London: Longman, pages 473–530.
- Boughton D (1994) *Evaluation and Assessment in Visual Arts Education*. Geelong, Australia: Deakin University Press.
- Boughton D (1996) Evaluating and assessing art education: issues and prospects. In: Boughton D, Eisner E and Ligtoet J (eds) *Evaluating and Assessing the Visual Arts in Education*. New York, USA: The Teachers College Press.
- Boughton D (1999) Framing art curriculum and assessment policies in diverse cultural settings. In: Mason R, Boughton D (eds) *Beyond Multicultural Art Education: International Perspectives*. New York, USA: Waxmann, pages 331–348.
- Boughton D, Eisner E, Ligtoet J (eds) (1996) *Evaluating and Assessing the Visual Arts in Education*. New York, USA: Teachers College Press.
- Bourdieu P, Passeron JC (1990) *Reproductions in Education Society and Culture*. London: Sage Publications.
- Broadfoot P (1996) *Education, Assessment and Society*. Milton Keynes: Open University Press.
- Eça T (2002) An evaluative study of studio art examinations at A/AS levels in Portugal and England. Unpublished PhD thesis for Centre of Art Education and International Research. Surrey: University of Roehampton.
- Eisner E W (1979a) *Educating Artistic Vision*. New York, USA: Macmillan.
- Eisner EW (1979b) *The Educational Imagination*. New York, USA: Macmillan
- Eisner E W (1985) *The Art of Educational Evaluation: A Personal View*. Philadelphia, USA: The Falmer Press.
- Eisner E W (1999) What do the arts teach? *Royal Society of Arts Journal* **2**: 42–51.
- EPPI-Centre (2002) *Review Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research. Version 0.9.5*. London: EPPI-Centre, Social Science Research Unit.
- Gardner H (1991) The assessment of student learning in the arts. Unpublished manuscript. Cambridge, Massachusetts, USA: Harvard University.

- Gipps C, Stobart G (1997) *Assessment: A teacher's guide to the issues*. London: Hodder and Stoughton.
- Harden A, Peersman G, Oliver S, Oakley A (1999) Identifying relevant primary research on electronic databases to inform decision-making in health promotion: The case of sexual health promotion. *Health Education Journal* **58**: 290–301.
- Hardy T (2002) AS level art: goodbye to the wow factor. *Journal of Art and Design Education* **21**: 52–59.
- Harlen W, Deakin Crick R (2002) A systematic review of the impact of summative assessment and tests on student motivation for learning. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- HMI (1989) *The Introduction of GCSE in Schools 1986-1988*. London: HMSO.
- Lindström L (1998) Criteria for assessing student performances in the visual arts. In: Piironen L (ed.) *Portfolio Assessment in Secondary Education and Final Examination. Report of EU Comenius 3.1 Project*. Helsinki, Finland: University of Art and Design Helsinki, Department of Art Education, pages 7–16.
- MacGregor RN (1990) Reflections on art assessment practices. *Journal of Art and Design Education* **9**: 315–328.
- Moon B, Shelton Mayes A (eds) (1994) *Teaching and Learning in the Secondary School*. London: Routledge and Open University.
- National Advisory Committee on Creative and Cultural Education (NACCCE) (1999) *All Our Futures, Creativity, Culture and Education*. London: DfEE/DCMS.
- Price E (1982) National criteria for a single system of examining at 16+ with reference to art and design. *Journal of Art and Design Education* **1**: 397–407.
- QCA (2004) *A Level Art and Design Review of Standards 1978–1998*. London: QCA.
- Rayment T (1999) Assessing national curriculum Art AT2, knowledge and understanding: a small-scale project at Key Stage 3. *Journal of Art and Design Education* **18**: 189–194.
- Sadler DR (1987) Specifying and promulgating achievement standards. *Oxford Review of Education* **13**: 191–209.
- Schönau D, Kárpárti A, Verhelst V (1996) *Towards a Standardized Examinations System, Arts Joint Hungarian-Dutch Project*. Arnhem: Cito and Hungarian Academy of Craft and Institute of Education.
- Schönau D (1998) Final examinations in the visual arts in the Netherlands. *Art Education* **47**: 34–49.
- Shohamy E (2001) *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Essex: Pearson Educational Ltd.

Skilbeck M (1985) A core curriculum for the common school. Based on Inaugural Lecture 2/6/1982, revised 1985. London: University of London Press.

Skilbeck M (1992) National Curriculum within the OECD. *Unicorn* **18**: 9–13.

Steers J (1988) Art and design in the national curriculum. *Journal of Art and Design Education* **7**: 303–323.

Steers J (1996) GNVQs: The context, the rhetoric and the reality. *Journal of Art and Design Education* **15**: 201–214.

Stobart G, Gipps C (1997) *Assessment: A Teacher's Guide to the Issues*. London: Hodder & Stoughton.

Wilson B (1997) Arts standards and fragmentations: a strategy for holistic assessment. *Arts Education Policy Review* **98**: 2–9.

Woods P, O'Shannessy J (2002) Re-introducing creativity: day 10 at Hackleton school. *Curriculum Journal* **13**: 163–182.

Zimmerman E (1992) Assessing students' progress and achievements in art. *Art Education* **45**: 14–24.

## Appendix 1.1: Advisory Group membership

### ***Review Advisory Group***

|                                       |  |
|---------------------------------------|--|
| Dr Dennis Atkinson                    | Lecturer Art Education, University of London<br>(Goldsmiths College) |
| Sheila Dowling                        | Senior Lecturer Art and Design, University College<br>Northampton    |
| Dr Richard Hickman                    | Senior Lecturer Art Education, University of<br>Cambridge            |
| Professor Rachel Mason (Director)     | Roehampton University  |
| Dr Trevor Rayment                     | Course Leader PGCE Art and Design, University of<br>Reading          |
| Dr Paul Rinne                         | Deputy Head and Head of Art, Sheen School                            |
| Dr Dorothy Bedford (Research Officer) | Roehampton University  |
| Dr John Steers (Co-Director)          | National Society for Education in Art and Design                     |
| Dr Sylvia Willerton                   | Education Consultant   |

### ***International members***

|                     |  |
|---------------------|--|
| Dr Doug Boughton    | Professor of Art Education, Northern Illinois University             |
| Dr Lee Emery        | Associate Professor, Arts Education, University of<br>Melbourne      |
| Dr Lynn Galbraith   | Associate Professor, Art Education, University of<br>Arizona, Tucson |
| Dr Folkert Haanstra | Professor of Art Education, Amsterdam School of the<br>Arts          |

### ***User members***

|                |   |
|----------------|---|
| Dan China      | Association of Advisors and Inspectors of Art and<br>Design |
| Melanie Fowke  | Art and Design teacher, St Guthlac's School, Spalding       |
| Dominic Harper | Art and Design teacher, Cowes Secondary School              |
| Michele Kitto  | Art and Design teacher, The Dragon School, Oxford           |
| Nick McKemy    | Art and Design teacher, Bacons College, London              |
| Vivienne Reiss | Arts Council England  |
| Robert Watts   | Primary teacher, Normand Park School, Fulham                |

## Appendix 2.1: Inclusion and exclusion criteria

### ***Inclusion criteria***

Subject: Art Education, Art, Design, Craft, Fine Art, Textiles, Photography, 3D, Drawing, Graphics, Painting, Print-making

Formal assessment: exams, tests, grading, coursework, portfolio, group work, projects, competence, skills, key skills, continuous assessment, evaluation

Level: secondary schools, grammar schools, comprehensive schools, independent schools, high schools, General Certificate of Education, General Certificate of Secondary Education, Scottish Certificate of Education, Certificate of Secondary Education, National Vocational Qualifications, O-level examinations, A-level examinations, AS-level examinations, Higher Grade Examinations, National Curriculum, Secondary School Curriculum, Key Stage 3, Key Stage 4

### ***Exclusion criteria***

Subject: Numeracy, Literacy, Performing, Arts, Dance, Music, State of the Art

Assessment: internal assessment, informal assessment

Level: primary education, further education, higher education, adult education, university education, gallery education, museum education

Other: pre-1977, not in English

**Table 2.1:** Exclusion criteria

|   |                                       |
|---|---------------------------------------|
| 1 | Not about topic – Art or Design       |
| 2 | Not about formal assessment           |
| 3 | Not about secondary- school education |
| 4 | Not empirical research                |
| 5 | Not in English                        |
| 6 | Pre-1977                              |

## Appendix 2.2: Search strategy

### Search sources

#### ***Educational research – specific sources***

Current Educational Research in UK (CERUK)  
Database of Education Systems in Europe (Standards) (EURYBASE)  
Education Information Management Exchange at NFER LEA Reports and Publications (EMIE)  
Education-line – BEI Service supporting educational research, policy and practice  
Information Network on Education in Europe (EUYDICE)  
International Review of Curriculum and Assessment frameworks (INCA)  
National Centre for Vocational Education Research (Australia) (NCVER)  
National Foundation for Educational Research (NFER)

#### ***Specialist research registers***

Digital Dissertations – online index to UK digital dissertations (not published)  
Dissertation Abstracts – online index to USA dissertations (not published)  
Educational Research Abstracts – online database  
Index to Theses – online index to UK dissertation titles (not published)

#### ***Bibliographic databases***

##### *Journal articles, conference papers, reports*

Australian Education Index (AEI)  
BIDS/ERIC  
British Education Index (BIDS/BEI)  
Electronic Table of Contents from British Library (ZETOC)  
General Ingenta Services (BIDS)  
International ERIC (includes BEI, AEI)  
Web of Science – Arts and Humanities  
Web of Science – Social Science Citation Index

##### *Journal databases*

EBSCO (Professional Development Collection)  
Ingenta e.g. JADE, Creativity Research Journal  
Ingenta Select (Catchword and Ingenta)

#### ***Online library catalogs***

##### *Books and journals – not specific articles*

Allison Research Index of Art and Design (ARIAD)  
Art, Design, Architecture and Media Information Gateway (ADAM)  
Arts Education Partnership – USA  
British Library Catalog  
Education-line – BEI Service supporting educational research, policy and practice  
M25 Library Catalogs – Educational Establishment libraries within M25

National Arts Education Research Centre – USA  
National Centre for Vocational Education Research (Australia) (NCVER)  
National Society for Education in Art and Design (NSEAD)  
Roehampton Library Catalog (OPAC)  
UK Higher Education Library Catalogs (UK HE OPACS)  
University Research Libraries plus British Library (COPAC)

***Education – specific sources***

Department for Education and Skills (DfES)  
EDEXCEL  
Institute of Education, University of London  
Learning and Skills Council (LSC)  
Office for Standards in Education (Ofsted)  
Qualifications and Curriculum Authority (QCA)

***Specialist research registers***

Digital Dissertations – online index to digital dissertations (not published)  
Dissertation Abstracts – online index to USA dissertations (not published)  
Educational Research Abstracts – online database  
Index to Theses – online index to UK dissertation titles (not published)

Table 2.2.1: Extract from prototype search log

| Source   | Date     | Search criteria  | Number of hits    | References downloaded | Saved as                                      | Import filter             | Notes                                      |
|--|----------|--|-------------------|-----------------------|---|---------------------------|--|
| BIDS/BEI   | 14/05/03 |  | 97                |                       | BEI1.txt                                      |                           |  |
| BIDS/BEI   | 16/05/03 |  | 45                |                       | BEI2.txt                                      |                           |  |
| BIDS/BEI   | 16/04/03 |  | 62 all duplicates | 0                     | BEI3.txt<br>BEI.enl                           |                           |  |
| BIDS/BEI   | 16/04/03 |  | 190               | 0                     | BEI4.txt<br>BEI.enl                           |                           |  |
| BIDS/BEI   | 16/04/03 |  | 41<br>26 BEI      |                       | BEI5.txt<br>BEI.enl                           |                           |  |
| BIDS/ERIC  | 14/05/03 |  | 1666              |                       |   |                           |  |
| BIDS General Ingenta Services  | 02/04/03 | Art and Education and Assessment                           | 30                | 17                    | Ingenta.txt                                   | Uncover (Ingenta)         | Some overlap with BIDS/BEI                 |
| EBSCO<br><a href="http://search.global.epnet.com/athens.asp">http://search.global.epnet.com/athens.asp</a> | 02/04/03 | Art and Education and Assessment                           | 182               | 131                   | EBSCOA.txt<br>EBSCOD.txt<br>BiblioManager Tab | AcadSearch Filter (EBSCO) | USR Journals link<br>Year is in date field |
| EBSCO  | 02/04/03 | Art and Education and Evaluation<br>Not Dance<br>Not Music | 147<br>31         | 116                   | EBSCO2.txt                                    |                           |  |
| COPAC  | 08/04/03 | Art and Education and Assessment                           | 53                | 6                     | COPAC1.txt<br>COPAC1.enl                      | COPAC.enz                 | Download via email                         |
| COPAC  | 11/04/03 | Art and Education and Assessment                           | 407               | 62                    | COPAC1.Txt<br>COPAC2.enl                      | COPAC.enz                 | Download via email                         |
| Index to Theses<br><a href="http://www.theses.com">www.theses.com</a>                                      | 02/04/03 | Art and Education and Assessment                           | 25                |                       | Can't Save Refs                               |                           | Web Page = Index to Theses Search.htm      |
| WOS Arts and Humanities<br><a href="http://wos.mimas.ac.uk">wos.mimas.ac.uk</a>                            | 02/04/03 | Art and Education and Assessment                           | 10                | 10                    | WOS.enl                                       |                           | Export directly to EndNote                 |
| British Library Journal Abstracts<br><a href="http://Zetoc.mimas.ac.uk">Zetoc.mimas.ac.uk</a>              | 16/04/03 | Art and Education and Assessment                           | 37                | 26                    | ZETOC1.txt<br>ZETOC.enl                       | Zetoc (MIMAS)             | Save Tagged Records Short Labels Format    |
| <a href="http://Zetoc.mimas.ac.uk">Zetoc.mimas.ac.uk</a>   | 16/04/03 | Art and Education and Evaluation                           | 16                | 9                     | ZETOC2.txt<br>ZETOC.enl                       |                           |  |

## Appendix 2.3: EPPI-Centre keyword sheet, including review-specific keywords

### V0.9.7 Bibliographic details and/or unique identifier

|  |   |   |  |
|--|---|---|--|
| <p><b>A1. Identification of report</b><br/>                 Citation<br/>                 Contact<br/>                 Handsearch<br/>                 Unknown<br/>                 Electronic database<br/>                 (Please specify.) .....</p> <p><b>A2. Status</b><br/>                 Published<br/>                 In press<br/>                 Unpublished</p> <p><b>A3. Linked reports</b><br/> <i>Is this report linked to one or more other reports in such a way that they also report the same study?</i></p> <p>Not linked<br/>                 Linked (Please provide bibliographical details and/or unique identifier.)<br/>                 .....<br/>                 .....<br/>                 .....</p> <p><b>A4. Language</b> (Please specify.)<br/>                 .....</p> <p><b>A5. In which country/countries was the study carried out?</b> (Please specify.)<br/>                 .....<br/>                 .....<br/>                 .....</p> | <p><b>A6. What is/are the topic focus/foci of the study?</b><br/>                 Assessment<br/>                 Classroom management<br/>                 Curriculum*<br/>                 Equal opportunities<br/>                 Methodology<br/>                 Organisation and management<br/>                 Policy<br/>                 Teacher careers<br/>                 Teaching and learning<br/>                 Other (Please specify.) .....</p> <p><b>A7. Curriculum</b><br/>                 Art<br/>                 Business studies<br/>                 Citizenship<br/>                 Cross-curricular<br/>                 Design and technology<br/>                 Environment<br/>                 General<br/>                 Geography<br/>                 Hidden<br/>                 History<br/>                 ICT<br/>                 Literacy – first language<br/>                 Literacy further languages<br/>                 Literature<br/>                 Maths<br/>                 Music<br/>                 PSE<br/>                 Physical education<br/>                 Religious education<br/>                 Science<br/>                 Vocational<br/>                 Other (Please specify.) .....</p> | <p><b>A8. Programme name</b> (Please specify.)<br/>                 .....</p> <p><b>A9. What is/are the population focus/foci of the study?</b><br/>                 Learners<br/>                 Senior management<br/>                 Teaching staff<br/>                 Non-teaching staff<br/>                 Other education practitioners<br/>                 Government<br/>                 Local education authority officers<br/>                 Parents<br/>                 Governors<br/>                 Other (Please specify.) .....</p> <p><b>A10. Age of learners</b> (years)<br/>                 0–4<br/>                 5–10<br/>                 11–16<br/>                 17–20<br/>                 21 and over</p> <p><b>A11. Sex of learners</b><br/>                 Female only<br/>                 Male only<br/>                 Mixed sex</p> | <p><b>A12. What is/are the educational setting(s) of the study?</b><br/>                 Community centre<br/>                 Correctional institution<br/>                 Government department<br/>                 Higher education institution<br/>                 Home<br/>                 Independent school<br/>                 Local education authority<br/>                 Nursery school<br/>                 Post-compulsory education institution<br/>                 Primary school<br/>                 Pupil referral unit<br/>                 Residential school<br/>                 Secondary school<br/>                 Special needs school<br/>                 Workplace<br/>                 Other educational setting (Please specify.) .....</p> <p><b>A13. Which type(s) of study does this report describe?</b><br/>                 A. Description<br/>                 B. Exploration of relationships<br/>                 C. Evaluation<br/>                     a. naturally-occurring<br/>                     b. researcher-manipulated<br/>                 D. Development of methodology<br/>                 E. Review<br/>                     a. Systematic review<br/>                     b. Other review</p> |
|--|---|---|--|

## Review-specific keywords

|  |  |
|--|--|
| A.1 Assessment context                                   | A.1.1 National<br>A.1.2 Regional/state<br>A.1.3 Local<br>A.1.4 Other (e.g. international)  |
| A.2 Assessment procedure                                 | A.2.1 Externally assessed<br>A.2.2 Externally moderated<br>A.2.3 Internally assessed<br>A.2.4 Internally moderated<br>A.2.5 Combinations (Please describe in 'details' box.)<br>A.2.6 Jury method<br>A.2.7 Collaborative assessment (teacher and student together) |
| A.3 Assessment component                                 | A.3.1 Coursework (e.g. portfolio, exhibition)<br>A.3.2 Terminal exam<br>A.3.3 Test<br>A.3.4 Combination of above<br>A.3.5 Other (please describe)  |
| A.4 Range of subject matter assessed                     | A.4.1 Practical<br>A.4.2 Theoretical (critical/analytical/historical etc.)   |
| A.5 Assessment marking                                   | A.5.1 Measuring scores (continuous numerical)<br>A.5.2 Grading according to categories (e.g. A, B, C, etc.)<br>A.5.3 Connoisseurship (qualitative)<br>A.5.4 Combination  |
| A.6 Assessment benchmarking                              | A.6.1 Criterion-referenced<br>A.6.2 Norm-referenced<br>A.6.3 Peer-referenced<br>A.6.4 Holistic<br>A.6.5 Other  |
| A.7 Discussion of the quality of assessment in the paper | A.7.1 Yes<br>A.7.2 No  |
| A.8 Impact on curriculum                                 | A.8.1 Yes<br>A.8.2 No  |
| A.9 Impact on teachers                                   | A.9.1 Yes<br>A.9.2 No  |
| A.10 Impact on students                                  | A.10.1 Yes<br>A.10.2 No  |
| A.11 Impact on student performance                       | A.11.1 Yes<br>A.11.2 No  |

## Appendix 3.1: Details of studies included in the review

**Table 3.1.1:** Review – findings and conclusions

| Study ID                     | Country                             | Study aims   | Study type  | Study design summary   | Findings and conclusions   |
|------------------------------|-------------------------------------|--|-------------|--|--|
| Atkinson (1998)              | England                             | To investigate the language teachers use in students assessment of drawings  | Description | Observation of teachers (1) assessing drawings of tables set for a meal by students aged 11=12 and (2) assessing drawings by GCSE pupils recorded in the form of a training video. No details are given of sampling, data collection or data-analysis techniques.  | Teachers' comments were predicated on a perspectival paradigm or highlighted formal properties. They valued and produce a particular kind of drawer. Students' drawing abilities were constructed, regulated and confined within specific practices and discourses.  |
| Bennet (1989)                | England                             | Teachers criterion-referenced assessment practices in General Certificate of Secondary Education examinations  | Description | A small group of teachers and GCSE students from mixed comprehensive schools were interviewed individually and the process of teachers assessing Year 11 students' art work for LEAG examinations was observed. Interview transcripts and researcher field notes were fed back to participants for additional comments. Few details of data collection and analysis are given. It is implicit that analysis was qualitative, involving researcher inferences and some triangulation.   | Teachers appeared ill-equipped to handle criterion referenced assessment. They experienced difficulties, using written grade descriptions to assess visual images and using exemplary photographs of students' artwork. Dialogue between teachers about how they are grading students' artwork is crucial to avoid orthodoxy. Exemplary visual materials are no substitute for looking at real works of art and may be prescriptive.   |
| Blaikie <i>et al.</i> (2003) | Canada, England and the Netherlands | 1.To determine whether students' experiences of portfolio assessment in art are different as a result of gender and national culture<br>2.To reveal what students value about portfolio assessment and if they see it as a valid preparation for their futures | Description | 17 high-school students from Britain, 29 from the Netherlands and 61 from Canada participated, of whom 64 were female and 26 were male. The study was based on self-completion survey questionnaire data, using a five-point Likert scale. The questions ranged from 'strongly agree' to 'strongly disagree', in the case of seeking students' opinions about assessment, and 'highly accurate' to 'inaccurate' in the case of students' actual experiences of assessment. The study also used statistics supplied by the English Qualifications and Curriculum Authority to compare the performance of males and females in the GCE A-Level Art and Design examination. | There was wide support for portfolio assessment and the opportunity to talk to teachers about their work. Students' lived experience of assessment and national assessment cultures influenced their preferences for certain kinds of pedagogy and assessment. Significant gender differences were that males were less likely than females to know and understand the qualities teachers looked for in their work; they considered it less important than females to know and understand criteria for assessment, engage in group critiques, and discuss their art with teachers. |

| Study ID                                     | Country              | Study aims   | Study type                         | Study design summary   | Findings and conclusions   |
|--|----------------------|--|------------------------------------|--|--|
| Eca (2004)                                   | England and Portugal | 1. To compare and evaluate Art and Design examinations at pre-university level in England and Portugal<br>2. To develop and test our new assessment instrument for pre-university art examinations in Portugal | Evaluation: Researcher-manipulated | The research had two stages. The first investigated the current situation in pre-university art examinations in Portugal and England. The second focused on development and evaluation of a new assessment instrument for Portugal. A survey questionnaire was used to establish art teachers and students' evaluations of the MTEP art examination and their needs in Portugal. Data about A-level examinations in England were collected through document analysis, observation of trials, interviews with teacher examiners and a small-scale questionnaire. The researcher was a participant observer in the trials of portfolio assessment instruments and procedures developed and tested out in five Portuguese schools in stage two. Inter- and intra-reliability of the new and existing instruments was compared using a Rasch-based rating scale to analyse data obtained from simulated marking of current examination scripts and the trial portfolios. | Serious reliability, validity and utility issues were identified in examinations in both counties and especially in Portugal. It was possible to increase validity and authenticity of assessment in Portugal by using portfolio-based tasks with no loss of reliability. However, a need was identified for InSET programmes for teachers. Positive outcomes included increased motivation in students and independent learning. Negative outcomes included an increased workload for teachers and students. Bias was a potential problem because of variations in access to resources and teacher commitment.  |
| MacGregor Lemerise, Potts and Roberts (1993) | Canada               | To describe Canadian high-school teachers 'methods of assessment in music, art and drama, which are modified by geography, experience and professional background'   | Descriptive: national survey       | The survey instruments were a self-completion questionnaire and semi-structured telephone interviews. School districts across Canada selected high schools and distribute the questionnaires to teachers. The main types of data collected were personal information about respondents; rankings of teachers' learning priorities in the arts; rankings of teachers' assignment priorities; frequency of use of assessment methods; use of types of assessment recording practices; comparison of provincial, school and class assessment; systems and the teachers feelings about assessment. The analysis itemised and compared data on participants, their methods and means of recording assessment and their comments. Teachers' gender, language, academic background and qualifications, level and length of teaching experience were compared against learning assignment priorities and methods, recording practices and feelings about assessment.         | Assessment was being used to compare students (normative assessment) and did not allow measurement of individual student achievement. Pressures exerted by parents and administrators as well as school districts affected the form and frequency of assessment. Assessment in the arts is not so personal that seeking a common basis for it across Canada would be a waste of time.<br><br>Criteria for assessment rated most highly were developing individuality and independence. Cumulative assessment of individual projects carried most weight and performance-oriented examinations were prioritised over written forms. Teachers were divided on the question of rewarding effort in arts classrooms. They preferred negotiated assessment to student self-evaluation. There was concern that assessment practices sometimes worked against the struggle for empathy in the student-teacher |

| Study ID       | Country | Study aims  | Study type | Study design summary  | Findings and conclusions   |
|----------------|---------|---|------------|---|--|
|                |         |   |            |   | relationship, and against socialisation.<br><br>Canadians do not give substantial coverage to history and criticism or emphasise consistent interpretation of assessment criteria by classroom teachers.   |
| Rayment (1999) | England | An investigation into the evidence needed to make legitimate comments/judgements about students' critical and contextual knowledge and understanding of Art and Design at Key Stage 3 (11-14 years)   | Evaluation | 21 Year 9 students were required to produce a series of self-portraits together with evidence of process through annotated worksheets and sketchbooks, and they were encouraged to reveal evidence of critical activity in their practical artwork. Concurrent reflective analysis of the project was carried out, evidence of process and product was annotated; and students self-assessments using common criteria in the form of the End of Key Stage 3 Statements (levels of performance) were negotiated with the researcher in individual interviews. The grades awarded by students and the researcher were compared with grades awarded by two art teacher colleagues and the results were analysed. | Students tended to underestimate their own achievements.<br><br>The assessment task provided a satisfactory method for making a formal assessment but not for assessing critical activity in routinely produced art work which values spontaneity and idiosyncrasy above formal evidence of critical engagement.<br><br>Teachers expressed doubts about whether the extra time and effort (utility) involved was justified. The model of critical activity was simplistic. Activities that prescribe task, process and criteria within limited parameters run the risk of destroying what they set out to assess. Student self-assessment contributes to formative evaluation.   |
| Sabol (1994)   | USA     | 1. To describe, examine and critique state level testing practices in visual arts<br>2. To consider the range and degree of agreement of item structure and content underlying curricula themes, concepts, visual arts vocabulary the relationship between content and structure of the tests and the related | Evaluation | 50 state departments submitted arts achievement test materials, test manuals and state art curriculum frameworks and guides for content analysis, and answered a self-completion questionnaire. Data related to the development and implementation of testing programmes were collected through telephone interviews with state art supervisors and curriculum consultants. The data were analysed quantitatively. The DBAE curriculum, which identifies aesthetics, art criticism, art history and production as basic disciplines of the visual arts, provided the conceptual framework for analysing curriculum content.   | Tests varied significantly in terms of structure and questions but a degree of agreement existed about content. The level of the tests suggested disagreement about appropriate structures and content needed for assessment of art learning. They did not adequately address the array of knowledge skills and processes found in state art curricula. The objective tests and production activities, which made up nearly half the components, provided only partial indications of student learning. Test items were not differentiated in terms of levels of learning and did not require the use of higher-level thinking or problem-solving skills. All except one required reading skills. Some states were considering more authentic assessment measures in the hope of creating programmes that more comprehensively measured student learning. Pressures exerted by |

| Study ID         | Country                                     | Study aims   | Study type                            | Study design summary   | Findings and conclusions   |
|------------------|---|--|---------------------------------------|--|--|
|                  |   | state curricula guides or frameworks   |                                       |  | parents and administrators, as well as school districts, affected the form and frequency of assessment.  |
| Willerton (2001) | England<br>Wales and<br>Northern<br>Ireland | To identify and evaluate the extent to which GNVQ Art and Design examinations fulfilled the vocational aims and the impact of increasing demands for external assessment | Evaluation:<br>naturally<br>occurring | <p>Willerton studied the impact of the examination on staff and students in her own school. Data were generated for six years from observation of practice, interviews, self-completion questionnaires and school/college analysis of school records and 30 case studies of students. Core information was obtained about course target group, resources, option patterns and styles of delivering vocational and key skills units, attendance industrial placements and progression routes into FE.</p> <p>Similar data was obtained from another 20 centres served by three awarding bodies and from questionnaires and interviews with 30 GNVQ staff and 60 students. Student and group achievements in the 21 centres were compared statistically, using national data about GCE A level student results.</p> <p>National statistics on student progression (GNVQ student entry distribution statistics, GNVQ student point-of-entry and distribution statistics, GCE A-level pass rates, and Art and Design Admissions registry [ADAR] higher education entry statistics) were analysed. A sample of student achievements at her own centre were compared with those of students undertaking GCE A-level.</p> | <p>Positive features of the qualification outweighed negative ones There was disparity between GNVQ courses in the 21 centres and variations in allocations of time and resources. Interpretations of procedures and the demands of external verifiers were inconsistent (over 30% of centres had non-specialist external verifiers). GNVQ had not proved to be as successful as the traditional GCE A level/foundation course route to entry into HE. There was little evidence of progression through GNVQ from Foundation GNVQ (Level 1), although there was evidence of successful progress from Intermediate GNVQ (Level 2) to Advanced GNVQ (Level 3) qualifications.</p> <p>A performance system that specifies discreet, unitised areas of competence is unsuited to Art and Design in which reflection on action is important. The increased emphasis in externally set tests undermined the basic philosophy of GNVQ as a student-centered approach to learning and ignored the interests of the target group students who under-achieved in more academic examinations.</p> |

Table 3.1.2: Review – research methods and results

| Study                        | Which methods are used to recruit people into the study?   | What was the total number of participants in the study (the actual sample)?                                       | Please describe the main types of data collected and specify if they were used (a) to define the sample; (b) to measure aspects of the sample as findings of the study?   | Which methods were used to collect the data?  | Which methods were used to analyse the data?  | How are the results of the study presented?   |
|------------------------------|--|---|---|---|---|---|
| Atkinson (1998)              | Not applicable (Please specify.)<br>The study is largely concerned with theoretical issues arising from assessment of children's drawings. | Not applicable (e.g. study of policies, documents, etc.)  | Details<br>There is no extraction of data as such; the paper is mainly descriptive/theoretical.   | Not stated/unclear (Please specify.)<br><br>Coding is based on:<br>Reviewers' inference   | Explicitly stated (Please specify.)<br>Analysis based upon the author's judgement with reference to specified theory  | Details<br>As a discussion  |
| Bennet (1989)                | Not stated/unclear (Please specify.)   | Not stated/unclear (Please specify.)<br>Refers only to a 'small group'  | Details<br>Inclusion of verbatim quotations suggests the interviews were tape-recorded and transcribed<br><br>(i) Interview transcripts<br>(ii) Research field notes<br>(iii) Teacher/student further comment notes   | One to one interview (face to face or by phone)<br>face to face in naturalistic setting<br><br>Type of interview not specified – may have been a combination<br>Observation<br>Researcher field notes | Not stated/unclear (Please specify.)<br>Implicit: qualitative involving researcher inferences and some triangulation  | Details<br>Throughout the text, including quotations from teachers                                      |
| Blaikie <i>et al.</i> (2003) | Not stated/unclear (Please specify.)   | Explicitly stated (Please specify.)<br>107 participants – 64 known to be female and 26 male (others unidentified) | Details<br>The data collected were based on a survey questionnaire, using a Likert scale. Participants were asked to respond to all questions by using the five-point scale encircling the number that most closely corresponded to their opinion about how portfolio assessment could be carried out most effectively and to encircle the number which | Self-completion questionnaire   | Not stated/unclear (Please specify.)<br>One would have to presume that as a Likert scale was used for the survey-questionnaire that the method used to analyse the data would have been one of counting the number of responses to each scaled question, converting the numbered responses into percentage responses and interpreting the data relative | Details<br>Figures with text.<br>Percentages of male/female agreement with the questionnaire statements |

| Study                          | Which methods are used to recruit people into the study?   | What was the total number of participants in the study (the actual sample)? | Please describe the main types of data collected and specify if they were used (a) to define the sample; (b) to measure aspects of the sample as findings of the study?   | Which methods were used to collect the data?   | Which methods were used to analyse the data?   | How are the results of the study presented?   |
|--------------------------------|--|---|---|--|--|---|
|                                |  |   | most accurately described their actual experiences of studying Art/Design and preparing to be assessed by their portfolio work.   |  | to the examination of the literature.  |   |
| Eca (2004)                     | Explicitly stated (Please specify.)<br>Teacher volunteers and students in their schools  | Explicitly stated (Please specify.)<br>175                                  | Details<br>Interviews<br>Questionnaires<br>Observation<br>Document analysis<br>Analysis of experimental examination   | Curriculum-based assessment<br>Focus group<br>One to one interview (face to face or by phone)<br>Observation<br>Self-completion questionnaire<br>Exams<br>Coding is based on: Authors' description   | Explicitly stated (Please specify.)<br>Quantitative and qualitative<br>Qualitative data obtained from observation, interviews and reports subject to descriptive analysis  | Details<br>Detailed description in PhD thesis   |
| MacGregor <i>et al.</i> (1993) | Explicitly stated (Please specify.)<br>The authors state: 'The Canadian Education Association Handbook (1993) was used to identify school districts of a size sufficient to suggest the existence within each district of at least one high school teaching art, music and drama.<br><br>'From the list, school districts were drawn to ensure representation...in every province and in | Explicitly stated (Please specify.)   | Details<br>Main types of data collected: personal information related to respondents; rankings of teachers' learning priorities in the arts; rankings of teachers' assignment priorities; frequency of use of assessment methods; use of types of assessment-recording practices; comparison of provincial, school and class assessment; systems and the teachers feelings about assessment<br><br>Responses to questionnaire were used to define the | One-to-one interview (face to face or by phone)<br>For the interviews, members of the research team had a series of questions to which informants might respond but structured them as they wished (semi-structured).<br>Self-completion questionnaire<br>Questionnaire title: Assessment in the Arts<br>The questions were organised around three bodies of information:<br>(1) background information about respondents (1-7)<br>(2) questions about aspects of assessment practices (8- | Implicit (Please specify.)<br>Quantitative data were analysed descriptively, with tabulations presented as frequencies and percentages. Qualitative data from comments in spaces were provided in the questionnaire and telephone interviews supplements the quantitative and is summarised and interpreted under six categories. Findings are grouped and compared (i) to answer the research questions and (ii) to discuss further issues for consideration. | Details<br>The statistical data from the questionnaire are presented in 27 tables. There are quotations from the interviews and questionnaire comments. |

| Study          | Which methods are used to recruit people into the study?   | What was the total number of participants in the study (the actual sample)?  | Please describe the main types of data collected and specify if they were used (a) to define the sample; (b) to measure aspects of the sample as findings of the study?  | Which methods were used to collect the data?   | Which methods were used to analyse the data?   | How are the results of the study presented?  |
|----------------|--|--|--|--|--|--|
|                | the Yukon and Northwest Territories. <sup>1</sup>  |  | sample for follow-up interviews  | 11)<br>(3) questions related to local or provincial requirements for assessment (questions 12–14)<br>They produced mainly quantitative data, but spaces were also provided for qualitative responses.  |  |  |
| Rayment (1999) | Not stated/unclear (Please specify.)<br>Captive group in a school environment  | Explicitly stated (Please specify.)<br>21 Year 9 students  | Details<br>Student interview responses<br>Student response sheets<br>Student annotated worksheets<br>Researcher field notes<br>Student tasks (self-portraits, worksheets and sketchbooks)<br>Student self-assessment<br>Teacher and researcher assessments of student work<br>Record of group discussion (students and researcher) | Curriculum-based assessment<br>Group interview<br>One-to-one interview (face to face or by phone)<br>Observation<br>Self-completion report or diary  | Explicitly stated (Please specify.)<br>'Reflective analysis' is not defined. Essentially data analysis was qualitative, but included analysis of inter-rater scoring of student work.  | Details<br>Discussion within the text but not a specific section or summary of conclusions   |
| Sabol (1994)   | Explicitly stated (Please specify.)<br>Art consultants or curriculum experts from the 50 states were contacted and asked to provide copies of visual arts achievement test materials, manuals, state curricula and answer a questionnaire. The | Not applicable (e.g. study of policies, documents, etc.)<br>The individuals consulted were not relevant to the study in that the focus was on the analysis of state art achievement tests. | Details<br>(1) Art assessment tests developed by each state<br>(2) Art curricula documentation (frameworks and guides) from each state.<br>(3) Questionnaire scripts completed by state departments<br>(4) Telephone interview scripts (between researcher and state departments)  | One to one interview (face to face or by phone)<br><br>Self-completion questionnaire<br>Questionnaires were sent to 50 state departments.<br>Other documentation<br>(1) Art assessment tests<br>(2) Art curriculum documentation – guides/frameworks | Explicitly stated (Please specify.)<br>Content-analysis methods recommended by Borg and Gall (1989), Merriam (1988) and Eisner (1981) were used. Data were analysed quantitatively. Absolute and relative frequencies were used where appropriate and terms such as mean, median, mode and | Details<br>The results were presented in statistical tables in the appendix section of the report. These findings were summarised in the conclusion section of the report. |

| Study            | Which methods are used to recruit people into the study? | What was the total number of participants in the study (the actual sample)? | Please describe the main types of data collected and specify if they were used (a) to define the sample; (b) to measure aspects of the sample as findings of the study?            | Which methods were used to collect the data?  | Which methods were used to analyse the data?         | How are the results of the study presented?                              |
|------------------|--|---|--|---|--|--|
|                  | consultants/experts were not the focus of the study      |   | All above used to measure aspects of the sample as findings of the study.  |   | percentages summarised content and central tendency. |  |
| Willerton (2001) | Not stated/unclear (Please specify)                      | Not applicable (e.g. study of policies, documents, etc.)                    | Details<br>The main type of data were qualitative, based upon 30n case studies of GNVQ students; observation of practice ('connoisseurship'); interviews; questionnaires; records. | One-to-one interview (face to face or by phone)<br>Observation<br>Self-completion questionnaire<br>School/college records (e.g. attendance records, etc.)<br>Secondary data, such as publicly available statistics<br>Other documentation<br>Coding is based on: Authors' description<br>Coding is based on: Reviewers' inference | Not stated/unclear (Please specify.)                 | Details<br>Incontinuous prose; the study was part of a PhD dissertation. |