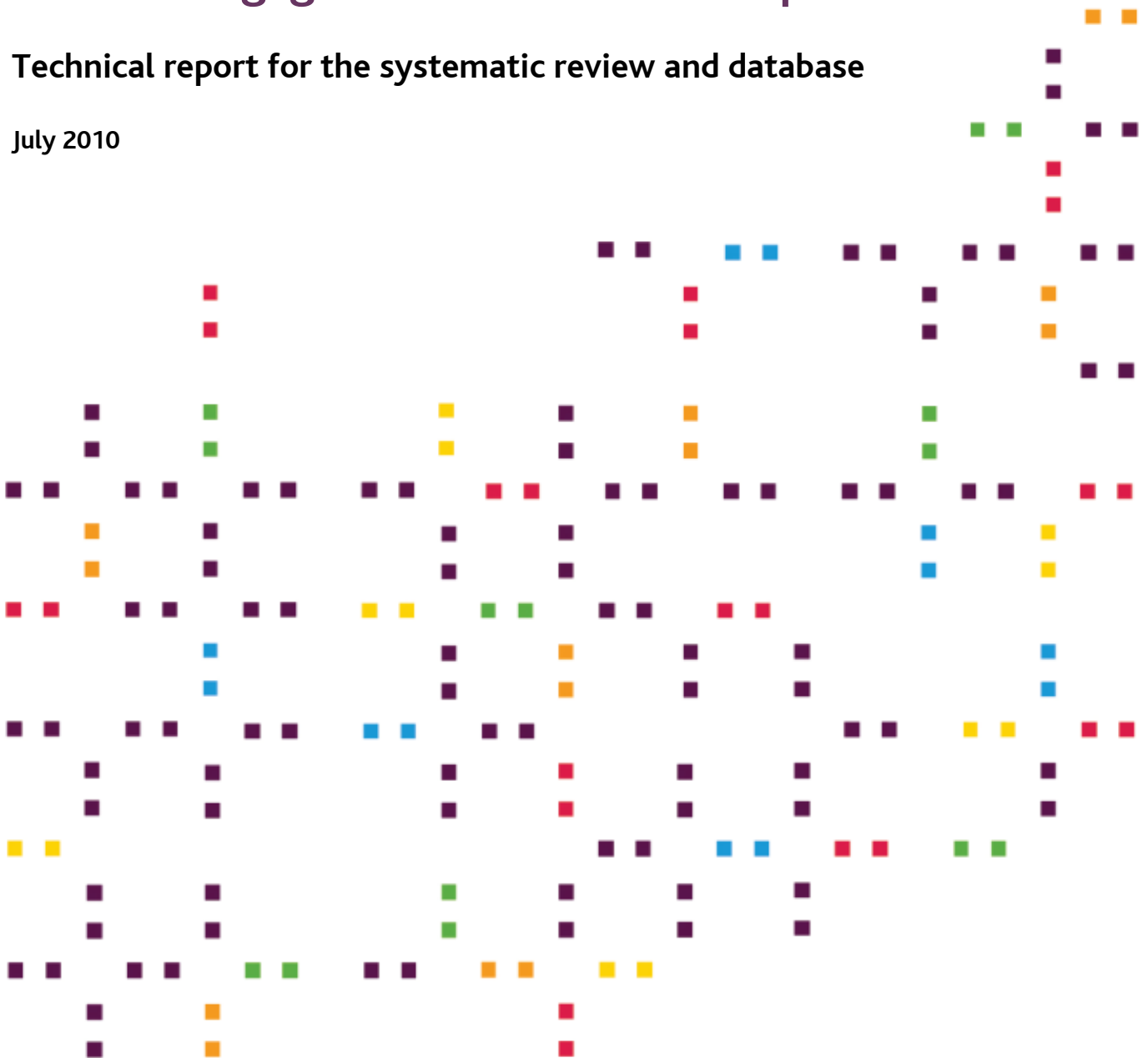


Understanding the drivers, impact and value of engagement in culture and sport

Technical report for the systematic review and database

July 2010



The CASE programme

The Culture and Sport Evidence (CASE) programme is a three-year joint programme of research led by the Department for Culture, Media and Sport (DCMS) in collaboration with the Arts Council England (ACE), English Heritage (EH), the Museums, Libraries and Archives Council (MLA) and Sport England (SE).

The work on this project was carried out by a consortium led by the EPPI centre with Matrix Knowledge Group

The EPPI-Centre (Institute of Education, University of London) are responsible for the content of this report:

Jan Tripney
Dr Mark Newman
Dr Karen Bird
James Thomas
Naira Kalra
Mukdarut Bangpan
Carol Vigurs



This report can be downloaded from the DCMS website:

http://www.culture.gov.uk/what_we_do/research_and_statistics/5698.aspx

Acknowledgements

We would like to acknowledge the support of Chloe Austerberry, Irene Kwan, Jeff Brunton, Sergio Graziosa, all of the EPPI-Centre, for information retrieval and technical support.

Contents

List of abbreviations	iv
1. Introduction	1
1.1. Background	1
1.2. Conceptual framework for the review	1
2. Methods used in the review	3
2.1 Outline of the chapter.....	3
2.2 General approach of the project.....	3
2.3 User involvement	3
2.4 Type of review	3
2.5 Review process	4
2.6 Stage 1: methods for creating a database of studies	6
2.7 Stage 2: mapping exercise methods	9
2.8 Stage 3: in-depth review methods.....	11
2.9 Deriving conclusions	16
3 Strengths and limitations of the review	18
3.1 Introduction.....	18
3.2 Searching and selection.....	18
3.3 Database	19
3.4 In-depth reviews	20
4. References.....	21
Appendices	22
Appendix 1: Case Board.....	22
Appendix 2: Search sources	23
Appendix 3: Search strings (general electronic databases).....	26
Appendix 4: Search strings (used to identify studies from pool of 12,439)	33
Appendix 5: Definitions of learning and social impacts.....	34
Appendix 6: Map coding tool.....	36
Appendix 7: Studies for which reports could not be obtained	43
Appendix 8: Identifying further studies through checking of relevant systematic reviews	44
Appendix 9: In-depth review data extraction and quality assessment tool.....	45
Appendix 10: Studies excluded from the arts participation synthesis	70
Appendix 11: Quality assurance (text mining)	71

List of abbreviations

ACE	Arts Council England
DCMS	Department for Culture, Media and Sport
EH	English Heritage
FE	Further education
MLA	Museums, Libraries and Archives Council
WoE	Weight of Evidence

1. Introduction

1.1. Background

The Culture and Sport Evidence (CASE) programme is a three-year joint programme of research led by the Department for Culture, Media and Sport (DCMS) in collaboration with the Arts Council England (ACE), English Heritage (EH), the Museums, Libraries and Archives Council (MLA) and Sport England (SE). The overall aim of the programme is to strengthen understanding of how best to deliver culture and sporting opportunities of the highest quality to the widest audience, generating the best outcomes for society.

As part of the CASE programme, DCMS commissioned the EPPI-Centre (Institute of Education, University of London) and the Matrix Knowledge Group to undertake a research project on the 'drivers and value of engagement with culture and sport'. The project used systematic review, analytical and statistical modelling techniques to begin the process of summarising what existing research evidence on specific areas can tell us about the impacts of engagement, and to use this and other data to understand why people engage in cultural and sporting activities, the benefits they obtain from it and the potential value, to them and to society as a whole, of that engagement. This report details the methods underlying the systematic review work. A statistical modelling exercise was conducted by Matrix Knowledge Group, the methods and results of which are reported elsewhere¹.

1.2. Conceptual framework for the review

In order to achieve the aims of the CASE programme, it was necessary to establish a coherent and consistent understanding of 'engagement in culture and sport'. To achieve this, an initial scoping project was undertaken to identify and outline the meanings and/or parameters of the key concepts: engagement, value, impact, sport and culture. The scoping project was an exploratory, iterative, cross-disciplinary exercise. Relevant documents were reviewed and a series of workshops with key sector stakeholders were held to gather their views and understandings of the key concepts. This information was utilised to support the development of the conceptual framework for the project.

The conceptual framework has four dimensions, which can be used flexibly and are open to different interpretations. The dimensions, and meanings attributed to them within the context of the 'drivers and value of engagement with culture and sport' project, are summarised in Table 1.1.

¹ See 'Understanding the drivers of engagement in culture and sport' and 'Understanding the value of engagement in culture and sport' summary or technical reports published alongside this report on the CASE website.

Table 1.1: Conceptual framework for the analysis

Dimension	Meaning
<i>Participant</i>	<ul style="list-style-type: none"> • The public (including, but not limited to: children; young people; 'excluded' groups, such as Black and Minority Ethnic people, people with disabilities, and communities/populations in particular geographical areas) • Cultural producers (including elites and professionals, as individuals or organisations)
<i>Activity/site</i>	<ul style="list-style-type: none"> • Sporting activities listed in the <i>Taking Part</i> survey² • Cultural (i.e., arts and heritage) activities/sites listed in the <i>Taking Part</i> survey • Museums, libraries and archives
<i>Outcomes</i> ³	Enjoyment, income, health, achievement, self-esteem, national pride/identity, international reputation, crime, government support, option-to-use, existence value, citizenship, social capital, understanding of UK culture, shared experience, escape, solace, inspiration, skills, community cohesion, community identify, excellence, productivity, investment

² Taking Part is a major, continuous survey of cultural and sport participation in England, commissioned by the Department for Culture, Media and Sport (DCMS) and its partner public bodies. Every year it collects information from several thousand (currently 15,000) adults aged 16 and over about their attendance at a wide variety of arts events, museums, galleries, libraries and heritage sites, and about their participation in creative activities and sport. The survey has been conducted since July 2005.

³ As no limits were placed on the type of outcome, this list is not exhaustive.

2. Methods used in the review

2.1 Outline of the chapter

This chapter describes the methods used in the systematic review. Section 2.2 explains the general approach of the project. Information about user involvement is detailed in section 2.3, followed by an outline of the type of review (section 2.4), the review process (section 2.5), methods used to create the database (section 2.6), and the methods used in the mapping exercise (2.7) and those used to undertake the in-depth review (section 2.8).

2.2 General approach of the project

The general approach to the project was informed by the aim of the CASE programme to produce robust and sustainable evidence to inform and steer public policy in culture and sport. Thus, the project as a whole was undertaken within a framework that 'evidence' should facilitate comparative analysis of (a) different drivers (including interventions) of people's engagement in culture and sport, and (b) the impact and value of engagement in different cultural and sporting activities (and indeed non-cultural/sporting activities). This, and the decision that any systematic reviews should where possible feed into the modelling exercise, formed the context in which decisions were made about the scope of evidence considered suitable for the review activities.

The CASE programme is very ambitious in its scope in a number of ways. Firstly, in its attempt to systematically utilise research evidence to inform decisions at a national level about how and where government can intervene most effectively to maximise value from engagement in culture and sport. Secondly, in the breadth of the scope of what is defined in policy terms as 'culture and sport', and, thirdly, in the desire to utilise 'new' research technologies to investigate and provide evidence to address these issues. Given the ambitious nature of the programme, the research activities undertaken as part of the 'drivers and value of engagement in culture and sport' project were in some senses exploratory and designed to provide a sustainable platform upon which further systematic investigation could be carried out at subsequent stages of the CASE programme.

There is little agreement across the culture and sports sectors about what constitutes 'impact', how or whether impact can be measured, how best to do so, and/or whether impacts demonstrated in one context can be generalised to another (Galloway, 2009). Whatever the different positions taken on these issues, every discussion of the issue, indeed the justification for any public policy on the topic, is that culture/sport is good for 'something', for 'somebody'. This is a causal claim, i.e., a claim that some kind of interaction with culture or sport produces an effect of some kind, whether it is immediate or long term, direct or indirect. The approach taken to the definition of high quality research in the review followed this logic, by selecting research that measured impacts quantitatively and attempted to establish cause and effect relationships between culture/sport and an outcome.

2.3 User involvement

CASE is governed by the CASE Board (see Appendix 1). Together, CASE Board members acted as an advisory group for this project and played a central role in the review process, for example, in establishing the conceptual scope of the review and choosing the focus of the in-depth review questions.

2.4 Type of review

The approach taken to reviewing the evidence in this project is different from the classic expert or narrative review usually undertaken in the social sciences, of which there are many examples in the fields of culture and sport. A systematic review is a piece of research that, like any other, follows an explicit method. It is a process for answering a review question that is driven by a pre-specified protocol. The review process includes the following steps:

- formulate the review question

- define the studies to be included with selection criteria
- search for studies (a systematic search strategy including multiple sources is used)
- screen studies for inclusion
- describe studies (keywording and/or in-depth data extraction).

At this stage a 'map' of studies in the field is produced that can be used to inform decisions about selecting a more specific focus for any in-depth review(s). In such a case, then additional inclusion criteria are applied to identify the relevant sub-set of studies. An in-depth review includes the following additional steps:

- in-depth data extraction
- assessing study quality (and relevance)
- synthesising findings.

Following this approach makes explicit any assumptions or biases in the review and also increases the potential value of the review, as it will be replicable, updateable and sustainable.

An additional feature of a systematic review is the incorporation of explicit quality assurance processes, including the use of trained staff, moderation, double coding, supervision and the use of specialist systematic review software.

The specific methods used at each stage of the systematic review carried out in the 'drivers and value of engagement in culture and sport' project are described in detail in the following sections.

2.5 Review process

The review was carried out in three stages:⁴

- Stage one: repository (database) of studies

The first stage of the review consisted of identifying all studies in the field of engagement in culture and sport. It produced a searchable database that is publicly available via the CASE website.

- Stage two: mapping exercise

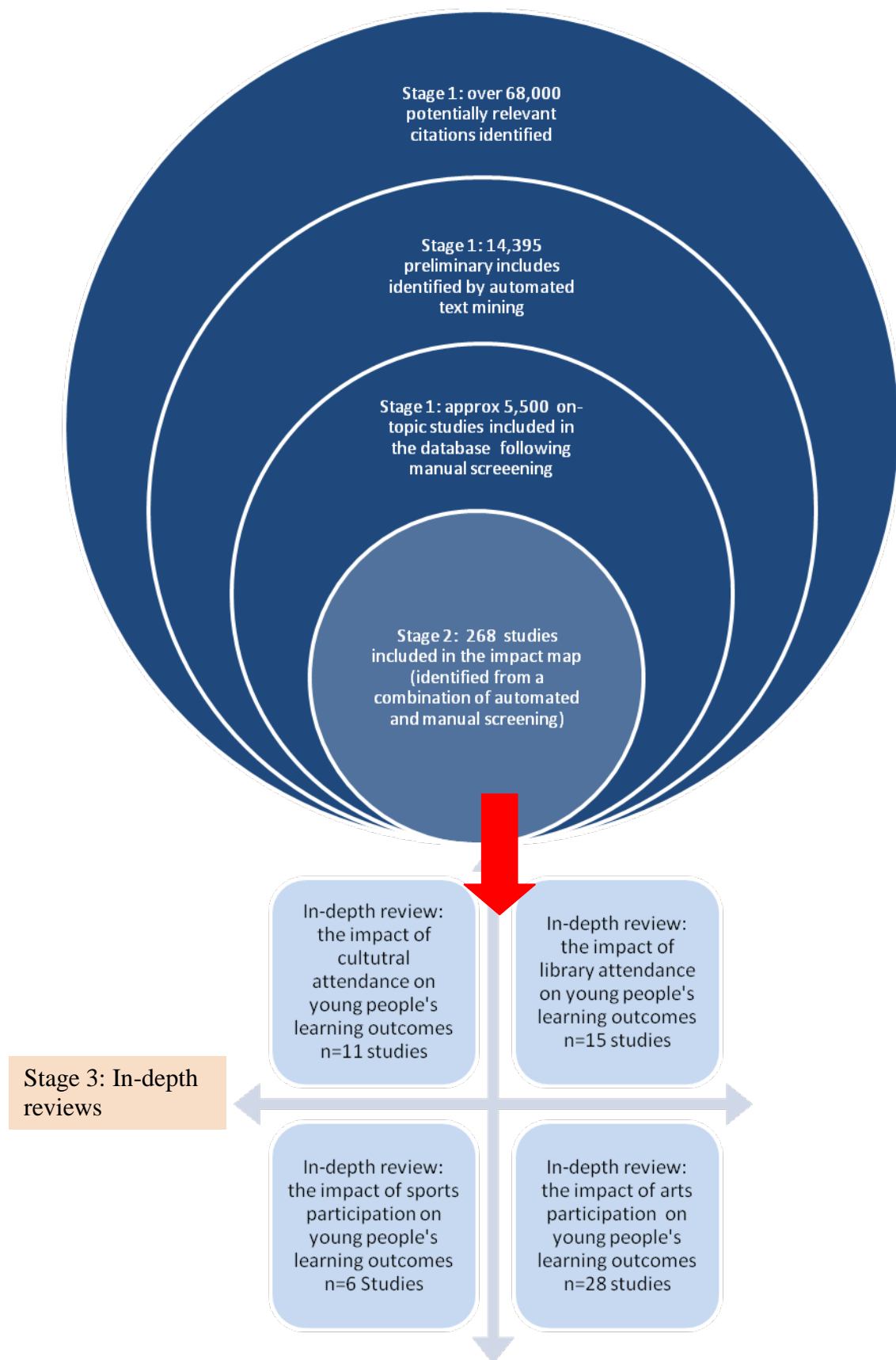
The second stage of the review involved describing (or mapping) a sub-section of the literature included in the database created at stage one. Quantitative impact studies were identified and relevant methodological and contextual information was collected. This information was presented in the form of a 'map' of research, which subsequently provided a basis for informed discussion and decision-making between the research team and the CASE Board about the focus of the in-depth reviews.

- Stage three: in-depth reviews (or synthesis) of a sub-set of studies

At the third (in-depth) stage of the review, a more detailed investigation of a focused sub-set of the wider literature was undertaken. This involved a synthesis of the findings of the selected studies, in order to provide answers to the in-depth review question.

⁴ These did not necessarily run sequentially.

Figure 1.1: summary of the flow of studies through the different stages of the review process



2.6 Stage 1: methods for creating a database of studies

2.6.1 Introduction

The first stage of the review consisted of identifying the research evidence about engagement in culture and sport. Reported in this section are the methods for selecting relevant studies for inclusion in the database (2.6.2), the search strategy (2.6.3), application of the selection criteria (2.6.4), and details of the quality assurance process (2.6.5).

2.6.2 Identifying relevant studies for inclusion in the database: inclusion/exclusion criteria

Selection criteria (inclusion and exclusion) defined the parameters of the review. These criteria were developed and agreed with the CASE Board (see Table 2.1). Studies were restricted to those published in English as there were insufficient resources for translation from other languages. The publication of *Use or Ornament? The Social Impact of Participation in the Arts* (Matarasso, 1997) and the formation of the Department of Culture, Media and Sport in the same year were identified as key chronological events for the field and a cut-off date of 1997 was agreed, with studies published prior to that date excluded from the review. Guided by the conceptual framework (section 1.2), all population groups, all outcomes and all engagement modes (attending, participating, deciding and producing) were within scope at this stage, but the range of cultural (arts and heritage) and sporting activities was limited to those included in the *Taking Part* survey. For a full list of included activities, see Appendix 6.

Table 2.1: Selection criteria (Stage 1: database creation)

No.	Exclusion criteria	Inclusion criteria
1.	<i>Study not published in English.</i>	<i>Study published in English.</i>
2.	<i>Study published prior to 1997.</i>	<i>Study published during or after 1997.</i>
3.	<i>Study is off-topic (i.e., is not about people's engagement in culture and/or sport).</i>	<i>Study is about engagement in culture and/or sport.</i>
4.	<i>Is not empirical primary research (e.g., opinion pieces, book reviews, bibliographies, news paper articles, editorials, strategy documents).</i>	<i>Is empirical primary research.</i>
5.	<i>Is solely methodological research.</i>	<i>Study is not solely methodological research (i.e., include if study involves a methodological aspect (e.g., the aim is to validate an instrument), but findings about people's engagement in culture and/or sport are also reported).</i>
6.	<i>Study is about engagement in cultural and/or sporting activities/sites, but not museums/libraries/archives or not as defined in the Taking Part survey (excludes, for example, studies about physical activity, walking, reading stories or poetry (as opposed to writing them), or architecture).</i>	<i>Study is about engagement in museums/libraries/archives, or cultural and/or sporting activities/sites as defined in the Taking Part survey.</i>
7.	<i>Study investigates a sports or exercise-based rehabilitation/treatment programme for people with pre-existing, non-chronic, physical health problems (for example, post-surgical interventions).</i>	<i>Study investigates a sports or exercise-based rehabilitation/treatment programme for people with pre-existing mental health problem or with chronic health problems, such as arthritis or back pain.</i>
8.	<i>Study only measures bio-medical outcomes (this criterion only applies to studies about engagement in sport).</i>	<i>Study measures any other outcome, in addition to bio-medical outcomes (this criterion only applies to studies about</i>

		<i>engagement in sport</i>).
--	--	-------------------------------

2.6.3 Identifying potentially relevant studies for inclusion in the database: search strategy

Using the broad (stage one) review question, the conceptual framework (see section 1.2) and the selection criteria, a number of possible searching strategies were identified by the research team for the CASE Board to consider. The six options presented to the Board varied on a number of parameters, according to the different emphasis placed on specificity and sensitivity. A comprehensive search strategy designed to locate as much literature as possible and to ensure thorough coverage of unpublished (grey) literature was selected. It was ambitious in scope, involving a large number and wide range of search sources.

Sources

The search strategy included eight main types of search source (see Appendix 2 for full details).

1. general bibliographic databases
2. specialist databases
3. journals not covered by the databases included in the search
4. websites of institutions known to have an interest in this area:
 - a) national and regional stakeholder organisations
 - b) UK research centres/departments/organisations
 - c) international research centres/departments/organisations
5. research funding bodies
6. subject specialists (publication lists)
7. search engines (Google/Google Scholar)
8. CASE Board.

As part of the piloting exercise to identify search terms, the reference lists of a limited number of reviews were checked and relevant studies were entered into the database, as were a number of relevant studies found serendipitously.

Search techniques

1. General bibliographic databases

Guided by the conceptual framework, search terms (or keywords) used to search the general bibliographic databases were developed iteratively.

- Relevant index and free-text terms were identified (both synonyms and antonyms) which could be used to describe the important concepts
- Pilot searches were undertaken to test the identified terms, which were then refined and used to search the bibliographic databases.

Full details of the search strings for individual databases are presented in Appendix 3. Where possible, database searches were restricted to literature published during the period 1997 to 2009.

2. Handsearching

Whilst conventional electronic database searches are very powerful, they are limited by the quality of indexing. 'Handsearching' was therefore undertaken for each of the remaining sources, to find relevant papers that had not been identified by the electronic searches. This involved carefully searching these sources on a 'page by page' basis and screening any potentially relevant items immediately against the exclusion/inclusion criteria. (Occasionally, relevant search terms – typically single terms such as 'sport' – were used to generate a number of hits which were then handsearched.) Any relevant items identified through handsearching were entered manually into the database.

All searches were conducted between June and August 2009. Citations identified in the above searches were imported into EPPI-Reviewer, the EPPI-Centre's specialist web-based systematic reviewing software (Thomas and Brunton, 2006).

2.6.4 Identification and selection of relevant studies for inclusion in the database: applying the inclusion/exclusion criteria

The broad scope of the review, in combination with a comprehensive search, generated a very high number of citation hits (over 68,000), many of which would be irrelevant. In a traditional systematic review, identification of relevant studies involves reviewers manually applying the pre-specified selection criteria to the titles and abstracts (and, where necessary, full reports) of papers identified using the search strategy. In this instance, within the resources available, it was not possible to manually screen all the potentially relevant citations (we estimated that this task alone would have taken at least 170 days, at a cost of approximately £70,000). Therefore, a different approach was adopted, involving the innovative use of text mining⁵ technology to support the process of study selection.

The identification and selection procedure followed by the reviewers was conducted in four phases:

1. First of all, reviewers manually identified studies that answered the broad review question. A random sample of 2,000 references was screened, and 243 (12%) were identified as meeting the inclusion criteria. Relevant items that had been identified as suitable for inclusion during handsearching (see section 2.6.3) were grouped with the 243 studies, bringing the total to around 1600. A quality assurance process was then carried out to check that these 1600 or so items covered the range of cultural and sporting activities and each of the four engagement modes (attending, participating, deciding and producing), as detailed in section 1.2. Attempts were made to fill any gaps by purposively searching for examples from across all four dimensions of the CASE framework. These steps resulted in a total of 1,733 included studies.
2. In the second phase, this initial sample of 1,733 studies was used to create an algorithm using the neural networks algorithm in SQL Server (MacLennan et al., 2009). This algorithm was used to classify the remaining studies as being relevant or not to the review. This process brought the number of references in the review down from the 68,432 identified by the initial search to 14,395 (12,662 plus 1,733 identified manually). The F-measure, which measures the accuracy of this process, was 0.77 (a score of 1 would indicate perfect classification).
3. Steps were taken to remove non-relevant items (using text clustering technology to identify non-relevant groups (or clusters) of items and then deleting the group and the studies therein). This process resulted in a total of 12,439 items which were provisionally included in the database.
4. With further quality assurance steps indicating that a significant proportion of non-relevant items still remained after the clustering exercise, the 12,439 items were manually screened against the inclusion/exclusion criteria. The main reason why these non-relevant studies remained was that they were non-empirical, which was difficult to detect using the algorithm. The process of manual screening reduced the number of included items to approximately 5,500.

The final phase in the development of the database ran concurrently with the mapping exercise (section 2.7) and the in-depth reviews (section 2.8).

A limited number of keywords were applied to the studies included in the database. For around half, the keywords were applied during the process of manual screening; for the remainder, an automated approach involving the use of 'search strings' to identify studies with particular characteristics was used.

⁵ Text mining searches for relevant documents based on the actual words that occur in the titles and/or abstracts, their frequencies, their proximity to other words, and so on.

2.6.5 Quality assurance process

This stage of the review followed EPPI-Centre procedures for maintaining quality. The search strategy was developed iteratively and tested using studies identified through handsearching. Extensive piloting of the inclusion/exclusion criteria took place. Members of the team who undertook screening took part in moderation exercises where samples of titles/abstracts were screened by everyone involved. Results were then compared and discussions held until any discrepancies were resolved. Independent audits of each team member's screening decisions were also carried out (on random samples of papers). These procedures aimed to ensure rigour and consistency between members of the review team in interpretation of the selection criteria. Quality assurance procedures were also built into the text mining process to increase the accuracy of this method of identifying relevant studies (see section 2.6.4). When the review was near completion, further measures were undertaken to assess the limitations of the searching and selection strategy (to be discussed in Chapter 3).

2.7 Stage 2: mapping exercise methods

2.7.1 Introduction

Within the resources available, it was not possible to describe (map) all studies that were identified as relevant to the review.⁶ Priority was therefore given by the CASE Board to understanding the impact of engagement in cultural and sporting activities. Such understanding underpins any analysis of the value of, and thus the case for, investment in this area. It was also agreed with the CASE Board that priority would be given to studies providing evidence of quantitative measures of impact on learning or social outcomes. In addition, the quality of the studies was taken into consideration. These four priorities (impact, quantitative, type of outcome, quality) guided the scope of the map that was produced. This map served as the basis for developing a more narrowly focused policy question to be answered by the final (in-depth) stage of the review.

The remainder of this section reports the methods for identifying (section 2.7.2) and describing (section 2.7.3) relevant studies for inclusion in the impact map, and details of the quality assurance process (section 2.7.4).

2.7.2 Identifying relevant studies for the impact map

The mapping exercise aimed to address the question:

What is the nature and extent of the available research literature on the quantitative impact on learning and/or social outcomes of engaging in cultural and/or sporting activities?

Answering a research question about impact requires establishing a cause–effect relationship. The scientific standards for inferring causation have been clearly established and the approach used in this review was to evaluate studies on the Maryland Scientific Methods Scale, a five-point scale for classifying the strength of a study design in relation to demonstrating causal relationships (Farrington et al., 2002). It assesses the quality of the design of studies with regard to minimising the effects of bias, and thus the scores generally reflect the level of confidence that can be placed in an evaluation's conclusions about cause and effect relationships, with a score of 5 (for randomised controlled experiments) indicating the strongest evidence.

In this review, high quality impact studies were therefore defined as those that had a control group and where, in the absence of random allocation to groups, both pre-test and post-test measurements of the outcome of interest were taken (i.e., expressed in terms of the Maryland Scientific Methods Scale, evaluations needed to score 3 or higher).

The initial plan was to include in the map only high quality studies (as defined above) that met the inclusion criteria. However, very few of these studies were identified for either the MLA or heritage sectors, and therefore for these sectors all quantitative studies of impact were mapped. The two groups of studies (high quality and lower quality) were identified separately,

⁶ Even before manual screening of the 12,439 items commenced, it was anticipated that the final number of included items would run into several thousands.

described below as part (a) and (b). In both instances, the procedure for identifying relevant studies involved a technological approach to narrow down the number of studies that reviewers had to screen manually.

Part a: identifying relevant high quality studies

Searching: Two search strings (see Appendix 4) were developed to identify items that were most likely to be high quality, quantitative studies (i.e., systematic reviews or primary research classified as level 3, 4 or 5 on the Maryland Methods Scientific Scale). Using these search strings, reviewers searched within the 12,439 studies that had provisionally been included in the database and identified 2,591 items.

Screening: A second set of selection criteria were developed (see Table 2.2) and used to manually screen these 2,591 items for relevance (i.e., to identify those that answered the impact question detailed at the beginning of this section). As indicated previously, the scope of the map was restricted to quantitative impact studies that had a control group and pre-test/post-test outcome measurements. The nature of the engagement mode also determined whether a study was included in the map. Two modes (deciding and producing) were considered less of an immediate priority and so were defined as out of scope. For studies about engagement in cultural activities, both the remaining modes (attending and participating) were considered relevant. However, for sport, only participating was of interest and studies about attending sporting events (either virtual or actual) were defined as out of scope. The definitions of what constituted 'learning' and 'social' impacts were pre-defined (see Appendix 5). These definitions were based on discussions in some of the relevant literature and the language/discourse used by stakeholders in the field, as captured in the workshop events during the scoping part of the project.

Table 2.2: Exclusion criteria (Stage 2: mapping exercise/high quality studies)

No.	Exclusion criteria*
1.	<i>Out of scope (as defined by the selection criteria outlined in Table 2.1)</i>
2.	<i>Not an impact study</i>
3.	<i>Non-systematic review</i>
4.	<i>No control group</i>
5.	<i>Not about participation (in sporting activities) or about attendance or participation (in cultural activities); i.e., exclude studies about the remaining two engagement modes – deciding and producing – or studies about attendance at sporting events (either virtual or actual)</i>
6.	<i>Not quantitative</i>
7.	<i>Not pre-test/post-test</i>
8.	<i>Non-relevant outcomes (e.g., health-related)</i>

*Only those studies excluded on no.1 would be removed from the database; all other criteria excluded items from the impact map only.

The application of these selection criteria identified 92 high quality primary research studies. This literature was coded for the map (see section 2.7.3).

Part b: identifying relevant lower quality studies

The process of coding the 92 studies revealed that there were few high quality studies in the MLA and/or heritage sectors. Following consultation with the CASE Board, it was agreed that for these two sectors all quantitative studies of impact would be mapped.

Searching: A further search string (see Appendix 4) was developed to identify all the remaining studies about museums, libraries, archives, and/or heritage (i.e., regardless of their quality or whether they were quantitative). This string was used to search within the original pool of studies (this now stood at 9,848: i.e., 12,439 minus the 2,591 high quality studies). This process identified 3,124 items.

Screening: These items were screened for relevance against a modified version of the exclusion criteria used to identify high quality impact studies (see Table 2.3). All outcomes (i.e., not solely learning or social) were relevant at this stage.

Table 2.3: Exclusion criteria (Stage 2: mapping exercise/lower quality studies)

No.	Exclusion criteria*
1.	<i>Out of scope (as defined by the selection criteria outlined in Table 2.1)</i>
2.	<i>Not impact study</i>
3.	<i>Not about museums, libraries, archives or heritage</i>
4.	<i>Non-systematic review</i>
5.	<i>Not about attendance or participation in museums, libraries, archives or heritage activities/sites; i.e., exclude studies about the remaining two engagement modes – deciding and producing.</i>
6.	<i>Not quantitative</i>

*Only those studies excluded on no.1 were removed from the database; all other criteria excluded items from the impact map only.

Application of these criteria resulted in an additional 116 (museums, libraries and archives) studies and 56 (heritage) studies being identified and coded for the map. (As these groups were not mutually exclusive, the total number of lower quality studies was 158.)

2.7.3 Describing studies in the impact map

The 250 included studies were coded using a framework developed specifically for this review in conjunction with the CASE Board. During this process, both contextual and methodological information was collected, focusing on areas of interest to the review. The coding exercise was primarily based on information contained in the study titles and abstracts, with details collected about:

- study design
- study participants
- engagement mode
- cultural/sporting activity/site
- outcomes measured.

Full details of the coding tool can be found in Appendix 6.

2.7.4 Quality assurance process

The mapping stage of the review followed EPPI-Centre procedures for maintaining quality. Piloting/moderation exercises were conducted to ensure consistency in interpretation and application of the selection criteria and map coding tool.

2.8 Stage 3: in-depth review methods

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

Taking account of the results of the map, policy priorities, and the resources and time available to complete the review, the focus for in-depth analysis was agreed with the CASE Board through a process of discussion and negotiation. A specific population group (children and young people) and a set of outcomes (learning) were selected as the foci of interest.

The in-depth review aimed to address the question:

What is the impact on children/young people's learning outcomes of engaging⁷ in cultural and/or sporting activities?

⁷ Engagement was defined as participation (in all four sectors) and attendance (in all sectors except sport). See exclusion criterion no. 5 in Tables 2.2 and 2.3.

To identify relevant studies for answering this question, the following steps were taken:

- Specific exclusion criteria were developed from the in-depth review question and applied to the 250 mapped studies (see Table 2.4). Two areas of related research were not considered within the scope of the in-depth review: studies that measured learning outcomes that were related to the cultural/sporting activity (criterion 3), and studies that examined changes to health-related knowledge/awareness (criterion 4). After application of the selection criteria, 48 studies were identified as answering the in-depth review question. Of these, six studies could not be included as we were unable to obtain copies of the reports (see Appendix 7).

Table 2.4: Exclusion criteria (in-depth review stage)

No.	Exclusion criteria
1.	<i>Study did not focus on children/young people</i>
2.	<i>Study did not measure at least one learning outcome</i>
3.	<i>Learning outcome measured was related to the cultural/sporting activity (e.g., sports performance, musical skills, etc.)</i>
4.	<i>Learning outcome measured was health-related knowledge/awareness (e.g., knowledge about STDs, HIV, etc.)</i>

- To identify further relevant high quality studies, the criteria in Table 2.4 were also applied to the nine systematic reviews that were identified during the mapping exercise and three were found to focus on children/young people and learning outcomes. These three reviews were used as a resource and their reference lists checked (see Appendix 8 for further details). For those primary studies listed that appeared relevant to the in-depth review, full reports were obtained to enable full-text screening against the inclusion criteria. In a number of cases, we could not obtain a copy of the full report and so could not assess eligibility. Seven additional high quality studies were included in the in-depth reviews.
- To identify further relevant studies for the MLA/heritage sectors, the references lists of four relevant non-systematic reviews were searched (see Appendix 8 for further details). For those primary studies that appeared relevant, full reports were obtained to enable full-text screening against the inclusion criteria. (Again, in a number of cases, we could not obtain a copy of the full report.) Seven additional studies were included. A further one relevant study was identified by the CASE Board.

At this stage, 57 studies were included for in-depth analysis. Later on, three further studies were identified as being relevant and also included (see section 3.2), bringing the final total to 60 studies.

2.8.2 Detailed description of studies in the in-depth review

Once selected, each included study underwent a process of in-depth data extraction in which the contents of each paper were summarised and evaluated according to pre-agreed categories. The tool used (see Appendix 9 for details) was developed in partnership with the CASE Board. Studies were coded on a number of variables, including: age group of participants; size of sample; type, duration and intensity of intervention; and outcomes measured. Information was also collected about the study context, study results for synthesis, and information on which to base judgements about the quality, trustworthiness and relevance of the study to the review.

Data extraction of each study's results focused on identifying the main quantitative measure of the impact of the intervention (i.e., engagement in cultural/sporting activities) on the learning outcomes of the participants (e.g., test scores).

2.8.3 Assessing study quality and weight of evidence for the review question

The quality of each included study was assessed using the EPPI–Centre's weight of evidence (WoE) framework. This has four components:

- **WoE A:** Assessment of the quality of the execution of the studies. In this review, the assessment focused on whether (1) knowledge of group allocation was concealed (blinding), (2) incomplete outcome data was addressed, (3) groups were treated equally, and (4) the outcome measures were reliable. Quality assessment was carried out at the level of an individual outcome, thus a study could have different quality scores for different outcomes. Studies were rated into five categories (high, medium/high, medium, low/medium, or low).
- **WoE B:** The appropriateness of the research design and type of analysis used for answering the review question. Firstly, the quality of the research design used in the included studies was assessed using the Maryland Scientific Methods Scale (MSMS). Secondly, for studies scoring level 4 or 5, two additional questions (generation of the allocation sequence and concealment of allocation) were applied. Studies were rated into the same five categories as used for WoE A.
- **WoE C:** The relevance of the study sample, measures, and actual analysis (or other indicator of focus of the study) to the review question. In this review, all studies were equally relevant to the in-depth review question. All had relevant outcome measures (learning outcomes) and comparable conceptual focus (aiming to examine the impact of engagement in culture/sport), and sampled the population of interest (children/young people). Therefore, all studies were rated high.
- **WOE D:** An overall weight of evidence, using a pre-established formula for moving from A, B and C to D. In this review, only A and B were taken into consideration and D could not be not greater than A or B (e.g., if A was medium/high and B was medium, then D would be medium). The higher the overall quality of the study, the more confident we can be that the result is a valid measure of the impact of the particular type of engagement in the particular cultural or sporting activities.

Table 2.5 outlines the procedure for calculating WoE.

Table 2.5: Study quality assessment framework

WoE A: Quality of the execution of the study	WoE B: Study design	WoE C: Study relevance	WoE D: Overall quality rating
<p>Was the knowledge of allocation to groups adequately prevented?</p> <p>Was incomplete outcome data addressed?</p> <p>Were the groups treated equally?</p> <p>Are the outcome measures reliable?</p> <p>Answers to individual questions were scored as follows and a total score calculated:</p> <ul style="list-style-type: none"> • yes = 1 	<p>What is the design of the study?</p> <p>Answers were scored as follows:</p> <ul style="list-style-type: none"> • Randomised controlled trial (MSMS level 5) = score of 5 • Well-matched⁸ comparison group pre-post test design (MSMS level 4) = score of 4 • Unmatched comparison group pre-post test design (MSMS level 3) = score of 		

⁸ Post-hoc statistical analysis used to control for differences between groups and comparison group considered to be well-matched to the intervention group on theoretically relevant factors (e.g., age, gender, etc).

<ul style="list-style-type: none"> • no = 2 • not relevant = 2 • unclear = 2 	<p>3</p> <ul style="list-style-type: none"> • Single-group pre-post test design OR comparison group post-test only design (MSMS level 2) = score of 2 • Single group post-test only design (MSMS level 1) = score of 1 <p>If studies scored 5 or 4, then two additional questions were applied.</p> <p>Was the allocation sequence adequately generated?</p> <p>Was the allocation adequately concealed?</p> <p>If the answer was 'no' to either of these questions, 1 was subtracted from the score for that study.</p>		
<p>High = score of 4 Medium/High = score of 5 Medium = score of 6 Low/Medium = score of 7 Low = score of 8</p>	<p>High = score of 5 Medium/High = score of 4 Medium = score of 3 Low/Medium = score of 2 Low = score of 1</p>	<p>Fixed at 'High'</p>	<p>Not higher than WoE A or WoE B</p>

2.8.4 Synthesis of evidence

Having coded the studies, the data were synthesised to bring together the studies which answered the in-depth review question. Before embarking on synthesis, a further selection criterion was applied to those studies rated at least 3 on the Maryland Scientific Methods Score. For these studies, inclusion was dependent on an overall (i.e., WoE D) quality score of at least medium. Application of this criterion resulted in four arts participation studies being excluded from further consideration (see Appendix 10 for details).

Four different in-depth reviews (syntheses) were undertaken, with sector and/or engagement mode determining the groups.

- Sport: participation
- Arts: participation
- Cultural (museums, galleries and heritage): attendance
- Libraries: attendance

Within these groups, the initial sweep of the individual studies identified that they were sometimes addressing slightly different sub-questions. Therefore, for each of the individual syntheses, the results are presented according to these sub-questions and outcomes.

The studies used a variety of approaches and methods to evaluate the relationship between engagement in cultural and/or sporting activities and children/young people's learning outcomes. The methods of synthesis used reflect the types of studies included in the in-depth reviews, and the detail and quality of reporting in these studies. The focus of each synthesis was on making comparisons across and between studies to identify what the overall pattern of results tells us about the answer to the question, rather than detailing the specific results of any individual study.

2.8.4.1 Synthesis methods for higher quality studies

A group of studies used research designs and methods of outcome measurement that allowed for a statistical approach to research synthesis. This approach involved conversion of the results data from each study into a common metric of 'effect sizes'. Effect sizes allow complex analysis to be presented in a way that is easy to compare and understand, and, as such, are an important tool in reporting and interpreting the effectiveness of a particular intervention relative to some comparison (Schagen and Elliot, 2004).

In this review, the effect size metric used was Hedges' g . This provides a standardised measure of the difference between the intervention group and control group in standard deviations. This metric is dimensionless and therefore allows comparison and combination across studies that use different outcome measures of the same conceptual phenomena (e.g., academic attainment). The effect size calculated represents the difference between the intervention and control groups in the baseline/post-intervention test 'change score'. The use of this effect size thus also helps to control for baseline differences between groups where random allocation was not used.

For most studies, effect sizes were calculated using data reported by the authors in the papers we had obtained. Where there was insufficient data to do so, or where an aspect of the study's findings was unclear, authors were contacted for further data and/or clarification. Details of which studies this applies to can be found in the relevant in-depth review reports (Bird et al., 2010; Tripney et al., 2010). Most studies reported multiple learning outcomes, and effect sizes were calculated for all outcomes, where possible. A number of different formulae were used to calculate effect sizes, depending on the data presented in the original papers.

The effect size provides an indication of the direction of the effect, with a positive effect size indicating that the participants who received the intervention had a better outcome than those in the control group. It also provides an estimate of the magnitude or size of the effect (although interpreting this requires further translation back into a dimension of some kind), and an estimate of the precision of the effect using confidence intervals. The confidence interval represents the range within which we can be 95% confident that the true result lies. Where a confidence interval crosses the line of no effect (zero in this case) we cannot exclude an effect in the opposite direction to that indicated by the point estimate.

Where the individual studies met the conditions required, meta-analytic techniques were used to synthesise the empirical research into an overall, or pooled, summary estimate of effect (Fitz-Gibbon, 1984). These conditions were that the studies were sufficiently similar that any pooled effect would provide a valid estimate of the impact of an intervention in a specific group for a particular phenomenon (notwithstanding any issues arising from the problems of bias that are not prevented by the study designs used in individual studies). Assessing this requires both statistical analysis and reviewer judgement, details of which are reported alongside the results of the meta-analysis in the relevant in-depth review reports (see '*Understanding the impact of engagement in culture and sport: a systematic review of the research on learning impacts for young people*' published alongside this report on the CASE website). Both fixed and random effects models were used to combine the individual study effect sizes to produce the pooled estimates. Each pooled estimate of effect provides a weighted average. Only one outcome from each study⁹ was included in any single meta-analysis, to avoid the same participant's outcomes being counted twice. Once the meta-analyses were completed, the studies were compared and explanations offered where effects varied systematically by variables such as intervention type, age of participants, etc.

2.8.4.2 Synthesis methods for lower quality studies

These studies were not suitable for statistical meta-analysis as they did not have suitable research designs and/or data. Studies in this category fell broadly into two groups. One type presented data in the form of frequencies based on questionnaire responses (i.e., a single group/post-test only design) with limited statistical analysis carried out by the study authors, so the results were not amenable to translation into a common metric of effect. For this type of study a narrative synthesis of the authors' reports of the results was undertaken.

⁹ In some studies several different cohorts of children were investigated and in such cases results from separate cohorts from the same study may be included in the same synthesis.

In the second type of study the results data used was the 'main' quantitative outcome measure of the impact of the intervention on the learning outcomes of the participants. The were either quantitative coefficients (or proportional differences) that measured the relationship between the independent variable (the intervention or control) and the dependent variable (academic attainment) and/or the authors narrative account of the results. The focus of the synthesis was on making comparisons across and between studies to identify what the overall pattern of results tells us about the answer to the question rather than the specific results of any individual study. The results of these studies are presented in the form of a coefficient that expresses the direction of the association. Either the slope is positive indicating that the intervention is associated with a better outcome or negative in which case the intervention is associated with lower outcome. Synthesis of these studies used a technique know as vote counting. The number of studies with positive or negative coefficients is simply added up and the balance either positive or negative is used to draw a conclusion. Unlike the primary study analysis statistical significance is not considered when deciding whether the findings of a study are positive or negative, as the aim is to identify what the overall body of studies indicates. Statistical significance reflects a number of different issues in study design including sample size, and the variables included in any model. Therefore it cannot be reliably interpreted across and between studies. Furthermore statistical significance does not provide any indication of practical importance or otherwise.

2.8.5 In-depth review: quality assurance process

The in-depth stage of the review followed EPPI-Centre procedures for maintaining quality. Piloting of the data extraction tool involved a percentage of studies included in the in-depth review being double coded (i.e., coded independently by two members of the team). The remainder of the studies were data-extracted by one reviewer and confirmed by a second. Discussions took place until any discrepancies were resolved. These measures were taken to develop and check consistency of data-extraction and quality assessment judgements between members of the review team.

2.9 Deriving conclusions

2.9.1 Introduction

A key problem in social science is how to determine which conditional statements (or conclusions) are scientifically interesting and which are pure speculation based on little more than the private intuitions of the person making the assertions. Unfortunately, there are no universally agreed firm criteria for making this judgement. The concern in the review context is to identify the likely degree of usefulness of an intervention as an instrument for achieving a particular outcome rather than its relative degree of absolute truth. From this perspective, the various findings from a review are neither absolutely true nor absolutely false, but rather have differing levels of warrant (Gordon, 1993). The warrant for conclusions from any study is based on a combination of the underpinning theory, the experiential knowledge of those involved, the research design and the quality of the research.

In the context of a systematic review of the kind reported here, the underpinning theory and the experiential knowledge are embedded in the design and execution of the interventions/studies included in the review, and in the design and conduct of the review itself. Thus, the review can be seen as a 'test' of the theories/knowledge embedded in each of the studies. The interpretation is derived from consideration of the result, i.e. the balance of evidence, the quality of the research design and the execution of the studies. The framework used to draw these interpretations is given in Table 2.6. It is an adaptation of the Maryland Scientific Methods Scale interpretation framework that takes into account study execution in addition to study quality (as in the quality assessment framework used for the review, outlined in section 2.8.3).

Table 2.6: Review interpretation framework

What works	<p>These programmes promote engagement in culture and/or sport.</p> <p>At least two evaluations of medium or greater quality with a positive pooled</p>
-------------------	---

	estimate of effect that excludes a result of no difference.
What does not work	At least two evaluations of medium or greater quality with a negative pooled estimate of effect that excludes a result of no difference.
What is promising	<p>These are programmes where the level of certainty from available evidence is too low to support generalisable conclusions, but where there is some empirical basis for predicting that further research could support such conclusions.</p> <p>Programmes are coded as promising if the pooled estimate of effect was found to be positive but did not exclude zero and there was at least one medium or better quality evaluation and the preponderance of the remaining evidence was also positive.</p>
What is unknown	Any programme not classified in one of the three above categories is defined as having unknown effect.

2.9.2 Interpreting effect sizes

A second aspect of interpretation concerns the interpretations of the standardised mean deviation used as the effect size in the meta-analysis. The effect size is dimensionless and as such requires translation back into a dimension which has meaning for policy and practice. Cohen proposes a general approach that labels an effect size of 0.2 to 0.3 a 'small' effect, around 0.5 a 'medium' effect and 0.8 to infinity, a 'large' effect (Cohen, 1988). The drawback of this approach is that it is context free, i.e., it takes no account of the importance of the effect, its likely applicability in the population, or the cost of achieving it. For example, for a school in which 50 per cent of pupils were previously gaining five or more A* – C grades, what using Cohen's system would be considered a 'small' effect of 0.2 would actually produce an increase from 50 per cent to 58 per cent, a difference that most schools would probably categorise as quite substantial (Coe, 2002).

However, it could also be argued that the important thing is how the effect size is derived. If it is derived from a high quality randomised experiment, then a difference of any size could be considered important. However, the quantification of extra physical properties, e.g., feelings or attitudes, cannot approach the level of measurement precision achieved in physical sciences (Nash, 2002). Thus, the issue of measurement error, and the fact that the instruments are measuring such latent variables, could account for many differences found in a study (Gorard, 2001).

3 Strengths and limitations of the review

3.1 Introduction

This systematic review undertaken as part of the 'drivers and value of engagement with culture and sport' project is, as far as we are aware, the first to have attempted to be comprehensive, systematic and transparent across such a wide body of literature in the field of culture and sport.¹⁰ As such, both the database and the individual in-depth reviews provide an important resource for the culture and sport communities, not only in their content but also in the development of systematic review methods for future investigation of the questions that remain unanswered in the field.

Due to the involvement of the culture and sport communities at the early stages of the review process, and the CASE Board throughout, the review has remained focused on policy- and practice-relevant questions and evidence.

The careful, detailed and explicit consideration given to the question addressed by each study, the quality of each study, and the quality assurance processes, mean that, whilst any reader might not agree with them, the basis for any conclusion reached in the review are clear and open to challenge.

However, this review represents only the first step in an ongoing process of building knowledge and understanding about the factors (including interventions) that drive engagement, and the impacts and values of that engagement. The in-depth reviews addressed only a very small part of the agenda of interest but importantly the systematic and comprehensive approach used means that it will be possible to utilise the database of studies to begin to address some of the other questions of interest in subsequent reviews.

3.2 Searching and selection

When conducting systematic reviews in social policy, it is difficult to identify and anticipate all of the terminology used to describe or explain the particular phenomenon of interest. When multiple phenomena are the subject of enquiry (as in this review) this problem is compounded. This is important as the terminology is used to create the search terms. Although our search strings were extensive, it is possible that not all relevant studies were identified.

It was only possible to conduct a review with such a broad scope in such a comparatively short space of time by using the new technologies of automated text mining. These reviews represent the first time, as far as we are aware, that such technology has been deployed in a systematic review. As such, there are no established rules or techniques for its optimal use. The review team worked in close collaboration with the text mining software team to develop the procedures in an iterative process of trial and error.

Using an automated approach to selection means that studies are selected for inclusion on the basis of using similar words and/or phrases as studies which are manually identified as being relevant to the review. Therefore, as the technology is dependent on the way that titles and abstracts are written, the automated approach may result in studies being included that are not in fact in scope. In this review, this was minimised by subsequent manual screening of all the studies selected for inclusion using the automated approach. More problematically, the use of text mining might also mean that studies that are within the scope of the review are not recognised as such by the software and are thus 'missed'.

Some quality checking was undertaken to assess the extent to which studies may have been 'missed', either in the initial search or by the text mining. Full details are reported in Appendix 11. We looked at the reference lists of four recent reviews in this area (covering each of the four sectors) to identify studies that, based on their title, would appear to have been relevant

¹⁰ There are systematic reviews on particular aspects of the field, such as factors influencing sports participation and impacts of the arts.

to this review and thus should have been in the database. The results of this exercise are shown in Table 3.1.

The results suggest that our initial search did not identify some relevant studies in each sector. A large proportion of the 'missed' studies were US dissertations, or other grey literature, which are not routinely included in electronic databases and which may have been obtained by the review authors' through personal contacts. Some 'missed' studies, however, were published in peer reviewed academic journals and it not yet clear why these were not detected by our search strategy. Ten relevant studies were not identified by the text mining, all of which were about some aspect of music participation. A likely explanation for this is that the studies, in their titles and/or abstracts, did not use a general descriptive term like 'music' but used a specific term like 'choir'. If the sample of included studies used to create the text mining search string did not include a study that used the specific term in its title or abstract then this term would not be in the text mining algorithm and thus any studies using only this term would not be identified as relevant to the review.

Table 3.1: Results of the quality assurance exercise

Sector	Study details	Relevant	Missed studies			Consequences
			Relevant studies (total) ¹	Missed studies (total)	Missed by search strategy	
Arts	Hallam (2009)	59	44	34	10	3
Sport	Long et al. (2009)	28	20	20	0	0
Libraries	BOP (2009)	18	8	8	0	0
Museums	BOP (2005)	15	1	1	0	0

1. Items in this column were identified by screening studies (based on title) in the reference lists of the four reviews against the database inclusion criteria.

2. Items in this column were identified by screening the missed studies against the in-depth review inclusion criteria. (This could only be completed for those studies for which we were able to obtain full texts; for nine studies, reports were unobtainable.)

All relevant items identified during the quality assurance exercise were manually entered into the database. The three available studies that were identified as addressing the in-depth review question were included in the synthesis of arts impact studies (see '*Understanding the impact of engagement in culture and sport: a systematic review of the research on learning impacts for young people*' published alongside this report on the CASE website).

3.3 Database

This project has created a publicly available database of studies in this field for the purpose of furthering understanding of past research and to facilitate and guide future inquiry. The database currently contains over 5,500 empirical studies, to our knowledge the largest of its kind in the world. Published and unpublished studies are included in the database, both UK-based studies and those conducted internationally. The literature represents a broad understanding of engagement in culture and sport; it is cross-sector and cross-disciplinary.

Designed to be updatable, the database offers substantial potential for the development of a cumulative knowledge bank in this field. As such, it makes an important contribution to the aim of the CASE programme to help co-ordinate and share culture and sport research at national, regional and local levels. However, given the size of the undertaking, only some keywords were applied manually; as a result, some error and/or omission are inevitable and may place some limits on the usage of the database. A further limitation is that included studies were restricted to those published in English.

When screening the 12,439 items against the database inclusion criteria, reviewers were unable to screen 19 items because there was insufficient information in the title and/or no abstract to make a decision. Two items passed the inclusion criteria for the database and on the basis of the titles/abstracts looked like relevant impact studies and so were requested from the library for full-text screening. Both reports were unobtainable and so their eligibility for the map and in-depth review could not be determined; however, the items are included in the database. Six items that were in the impact map and met the inclusion criteria for the in-depth review (on the basis of the abstract) did not arrive in time to be full-text screened for inclusion in the in-depth review. These six items are also included in the database.

3.4 In-depth reviews

There were fewer than anticipated high quality studies of the impact of engagement in culture and sport on children/young people's learning outcomes. Even with the addition of the six studies that we were unable to obtain (see Appendix 7), the overall yield of studies remains small. With one exception, all of the higher quality studies identified were of arts or sports participation, but even in these areas, relatively few studies were identified. Given the comprehensive scope of the review and the extensive searching strategy, it seems unlikely that many studies have been missed. The comparatively small number of studies identified is, however, not particularly unusual in social policy systematic reviews. To some extent, the apparently small number reflects the specificity of the in-depth review questions. The use of quantitative evaluation approaches is also contested in the field (see Galloway, 2009) and thus the requirement that studies used quantitative measures and experimental designs will also have affected the yield.

The sports and arts participation in-depth reviews included only studies that were graded as medium quality and above. There were approximately 170 quantitative impact studies on arts or sports engagement that did not meet this quality threshold. Given the difficulty in making quality judgements, it is possible that some of these studies were 'as good as' some of the included studies. Furthermore, these studies may provide additional understanding of the impact of a particular intervention where high quality studies are already identified. A useful extension to the review might be to look at the lower quality arts and sports studies and see where they fit with the available higher quality evidence.

The amount and quality of evidence available about cultural attendance (galleries, MLA and heritage sectors) is very limited. This may, in part, be due to the limitations of adopting the *Taking Part* survey definitions of cultural and sporting activities/sites in this review. However, even if the definitions, of heritage in particular, were wider, it is felt unlikely that many high quality quantitative studies would have been identified, as quantitative approaches to the evaluation of impact using experimental designs and measured outcomes have not traditionally been used in these sectors.

4. References

- Burns Owens Partnership Consulting (2009) *Capturing the impact of libraries: final report*. London: Department for Culture, Media and Sport.
- Burns Owens Partnership (2005) *New directions in social policy: developing the evidence base for museums, libraries and archives in England*. London: Museums, Libraries and Archives Council.
- Coe R (2002) It's the effect size, stupid: what effect size is and why it is important. Paper presented at: *British Educational Research Association Annual Conference*, Exeter, 12-14 September.
- Cohen J (1988) *Statistical power analysis for the behavioral sciences* (second edition). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Farrington D, Gottfredson D, Sherman L, Welsh B (2002) The Maryland Scientific Methods Scale. In Farrington D, MacKenzie D, Sherman L, Welsh L. (eds) *Evidence based crime prevention*. London: Routledge, pages 13-21.
- Fitz-Gibbon CT (1984) Meta-analysis: an explication. *British Educational Research Journal* 10(2): 135-144.
- Galloway S (2009) Theory-based evaluation and the social impact of the arts. *Cultural Trends* 18(2): 125-148.
- Gorard S (2001) *Quantitative methods in educational research: the role of numbers made easy*. London: Continuum.
- Gordon S (1993) *The history and philosophy of social science*. London: Routledge.
- Greenhalgh T, Peacock R (2005) Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *British Medical Journal* 331: 1064-1065.
- Hallam S (2009) *The power of music: its impact on the intellectual, social and personal development of children and young people*. Unpublished report.
http://www.ioe.ac.uk/Year_of_Music.pdf (accessed 18 January 2010).
- Long J, Hylton K, Spracklen K, Ratna A, Bailey S (2009) *A systematic review of the literature on black and minority ethnic communities in sport and physical recreation*. London: Sport England.
- MacLennan J, Tang Z, Crivat B (2009) *Data mining with Microsoft SQL Server 2008*. Indianapolis, IN: Wiley.
- Matarasso F (1997) *Use or ornament? The social impact of participation in the arts*. Stroud: Comedia.
- Nash R (2002) A realist scheme for social explanation: on 'numbers and narratives', *Building Research Capacity*, 4: 1-4.
- Schagen I, Elliot K (eds) (2004) *But what does it mean? The use of effect sizes in educational research*. Slough: National Foundation for Educational Research.
- Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C (2000) Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Controlled Clinical Trials* 21: 476-487.
- Thomas J, Brunton J (2006) *EPPI-Reviewer 3.0: analysis and management of data for research synthesis*. EPPI-Centre software. London: Social Science Research Unit, Institute of Education, University of London.

Appendices

Appendix 1: Case Board

Senior Responsible Officer: Anita Charlesworth, DCMS, Chief Analyst

Programme Manager: Adam C Cooper, DCMS, Head of Research

Vivienne Avery, DCMS, Chief Statistician

Harman Sagger, DCMS, Economic Adviser

Catherine Bunting, ACE, Director of Research

Laura Clayton, EH, Head of Social and Economic Research

Ailbhe McNabola, MLA, Head of Research

Nick Rowe, Sport England, Head of Research

Appendix 2: Search sources

The following sources formed the basis of the search strategy used to identify relevant literature for inclusion in the database.

1. *General Bibliographic Databases*

- ASSIA (Applied Social Sciences Index and Abstracts)
- SSCI (Social Science Citation Index) and AHCI (Arts and Humanities Citation Index)
- ERIC (Education Resources Information Centre)
- Medline
- BHI (British Humanities Index)
- IBSS (International Bibliography of the Social Sciences)
- PsycInfo
- Econlit

2. *Specialist Databases*

- Impact database
- Museums Libraries Archives Research and Evaluation Database
- Arts Research Digest

3. *Specialist Journals*

A selection of journals not covered by the general bibliographic databases listed in (1) was searched.

- International Review for the Sociology of Sport
- Sport in Society
- Engage Journal
- Cultural Trends
- Visual Culture in Britain

4. *Websites*

(a) National and Regional Stakeholder Organisations

- Department for Culture, Media and Sport
- Sport England
- Arts Council England
- English Heritage
- UK Sport
- Youth Sports Trust
- Department of Health
- Central Council for Physical Education
- Big Lottery Fund
- Fitness Industry Association
- Women's Sport and Fitness Foundation
- Sporting Equals
- English Federation of Disability Sport (EFDS)

- Commission for Architecture and the Built Environment (CABE)
 - Heritage Lottery Fund (HLF)
 - Creative and Cultural Skills
 - SkillsActive
 - Sports Coach UK
 - Craft Council
 - Communities and Local Government (CLG)
 - Her Majesty's Treasury (HMT)
 - Cabinet Office
 - Ofsted
 - DEMOS
 - National Audit Office
 - Audit Commission
 - Local Government Association (LGA)
 - Leisure Studies Association
 - National Foundation for Educational Research
- (b) UK Research Centres/Departments/Organisations
- CultureMap London
 - ESRC Centre for Research on Socio-Cultural Change (CRESC)
 - Loughborough University: Institute of Sport and Leisure Policy
 - Scottish Government: Culture, External Affairs and Tourism Research Network
 - Sheffield Hallam University: Sport Industry Research Centre
 - University of Chester: Chester Centre for Research into Sport and Society
 - University of Glasgow: Centre for Cultural Policy Research
 - University of Leicester: Research Centre for Museums and Galleries
 - University of Newcastle upon Tyne: International Centre for Cultural and Heritage Studies
 - University of Warwick: Centre for Cultural Policy Studies
- (c) International Research Centres/Departments/Organisations
- National Endowment for the Arts (US)
 - North American Society for the Sociology of Sport (US)
 - The Social Impact of the Arts Project (research centre at the University of Pennsylvania School of Social Policy and Practice) (US)
 - Canadian Council for the Arts (Canada)
 - Canadian Heritage (Canada)
 - Australian Council for the Arts (Australia)
 - Australian Sports Commission (Australia)
 - Council of Europe: Cultural Policy Research (Europe)
 - European Commission: Sport (Europe)

- European Cultural Foundation
5. *Social science research funding bodies*
- ESRC Society Today
 - Arts and Humanities Research Council
6. *Subject specialists*

Publication lists for specific individuals were searched (using university profile web pages or a Google Scholar search) to see if they contained relevant literature that had not previously been identified.

- Janet Ruiz
- Susan Galloway
- Christine Hamilton (Christine Hamilton Consulting)
- Adrienne Scullion (Scottish Executive Social Research)
- Eleonora Belfiore (Warwick University)
- Fred Coulter (Stirling University)
- Sara Selwood (City University),
- Javier Stanziola (Leeds University)
- Mike Savage (Manchester University)
- Andy Miles

7. *Google/Google Scholar*

As part of the piloting exercise to identify search terms, the reference lists of a limited number of literature reviews (systematic and non-systematic) reviews were checked. A limited, but focused, search for relevant reviews was carried out using Google/Google. Relevant items were entered into the database.

To identify relevant primary studies, a search string was entered into Google /Google Scholar and the top 50 hits were screened.

8. *CASE Board*

The CASE Board made requests to relevant stakeholders for information about studies that fell within the scope of the project and forwarded details to the review team. These items were then screened for inclusion in the review.

A number of relevant studies were found serendipitously and these too were added to the database.

Appendix 3: Search strings (general electronic databases)

Applied Social Sciences Index and Abstracts (ASSIA)

((((KW=(engage* or participat* or visit* or access* or aware* or join* or volunteer* or attend* or watch* or listen* or perform* or rehears* or creat* or member* or campaign* or archive* or collect* or produc*)) or(KW=((go to) or (going to) or (take part) or (taking part) or (decision making))) or(KW=(effect* OR impact* OR value* OR benefit* OR implication* OR advantage* OR disadvantage OR disadvantages or factor* or reason* or barrier* or facilitator*)))

and

((DE=("sports" or "archery" or "athletics" or "baseball" or "basketball" or "wheelchair basketball" or "bat and ball games" or "bowls" or "boxing" or "climbing" or "rock climbing" or "abseiling" or "competitive sports" or "cricket" or "endurance sports" or "extreme sports" or "figure skating" or "football" or "american football" or "rugby" or "quad rugby" or "golf" or "miniature golf" or "gymnastics" or "acrobatics" or "handball" or "hang gliding" or "hockey" or "ice hockey" or "ice skating" or "international sports" or "martial arts" or "ju jitsu" or "karate" or "tai chi" or "mountaineering" or "netball" or "olympic games" or "orientteering" or "racing" or "cycle racing" or "horse racing" or "rollerblading" or "running" or "skiing" or "snowboarding" or "squash" or "team sports" or "tennis" or "volleyball" or "watersports" or "kayaking" or "sea kayaking" or "rowing" or "sailing" or "swimming" or "white water rafting" or "windsurfing" or "wrestling") or(DE="culture") or(DE="arts") or(DE=("art" or "aboriginal art" or "contemporary art" or "drawings" or "cartoons" or "computer drawings" or "figure drawings" or "line drawings" or "paintings" or "impressionistic paintings" or "post impressionistic paintings" or "nature paintings" or "nude paintings" or "portraits" or "selfportraits" or "postmodern art" or "public art" or "ritual art" or "street art") or(DE="plays") or(DE=("films" or "documentary films" or "educational films" or "erotic films" or "gangster films" or "horror films" or "silent films" or "suspense films" or "war films" or "western films")) or(DE="literature") or(DE="concerts")) or(DE=("culture" or "popular culture" or "traditions")) or (DE=("museums" or "archives" or "libraries" or "heritage"))))

OR

((((AB=(sport* OR swimming OR diving OR cycling OR BMX OR cyclo-cross OR biking OR bowls OR bowling OR aerobics OR gym OR judo OR karate OR taekwondo OR taekwondo OR self-defence OR tai chi OR weight training OR body building OR weightlifting OR gymnastics OR snooker OR pool OR billiards OR darts OR rugby OR football OR camogie OR hurling OR handball OR shinty OR cricket OR hockey OR archery OR baseball OR softball OR netball OR tennis OR badminton OR squash OR basketball OR athletics OR jogging OR cross-country OR running OR angling OR fishing OR yachting OR sailing OR canoeing OR windsurfing OR boardsailing OR skating OR curling OR golf OR skiing OR horse riding OR climbing OR mountaineering OR trekking OR shooting OR volleyball OR orientteering OR rounders OR rowing OR triathlon OR boxing OR waterskiing OR lacrosse OR fencing OR yoga)) OR AB=(dance exercise) OR (keep fit) OR (motor sports) OR (pitch and putt) OR (tae kwon do) OR (martial arts))) or(AB=("museum" or "museums" or "gallery" or "galleries" or "library" or "libraries" or "archive" or "archives" or "heritage")) or(AB=((historic* city) or (historic* town) or (historic* building) or (historic* park) or (historic* garden) or (historic* landscape) or (historic* transport system) or (historic* place of worship) or (archaeological site) or (heritage site) or (historic* interest) or (historic* place) or (historic* space) or (historic* environment) or (historic* site) or (furniture making) or (jewellery making))) or(AB=(monument or castle* or ruin* or photography or sculpture* or video* or festival* or drama* or theat* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or painting* or drawing* or printmaking or animation* or textile* or craft* or calligraphy or pottery or poetry)))

near

((AB=(engage* or participat* or visit* or access* or aware* or join* or volunteer* or attend* or watch* or listen* or perform* or rehears* or play* or writ* or make or making or makes or creat* or buy* or member* or campaign* or teach or teaching or instruct* or train* or archive* or document* or collect* or produc*)) or (AB=((go to) or (going to) or (take part) or (taking part) or (decision making))) or(AB=(effect* OR impact* OR value* OR benefit* OR implication*

OR advantage* OR disadvantage OR disadvantages or factor* or reason* or barrier* or facilitator*))))

International Bibliography of the Social Sciences (IBSS)

((TX "Keep Fit" OR TX "aerobics" OR TX "swimming" OR TX "diving" OR TX "cycling" OR TX "biking" OR TX "bowls" OR TX "bowling" OR TX "gym" OR TX "judo" OR TX "karate" OR TX "taekwondo" OR TX "tai chi" OR TX "weight training" OR TX "weightlifting" OR TX "body building" OR TX "gymnastics" OR TX "snooker" OR TX "pool" OR TX "billiards" OR TX "darts" OR TX "rugby" OR TX "camogie" OR TX "hurling" OR TX "handball" OR TX "shinty" OR TX "cricket" OR TX "hockey" OR TX "softball" OR TX "netball" OR TX "tennis" OR TX "badminton" OR TX "squash" OR TX "basketball" OR TX "athletics" OR TX "jogging" OR TX "cross-country" OR TX "running" OR TX "yachting" OR TX "sailing" OR TX "canoeing" OR TX "windsurfing" OR TX "skating" OR TX "curling" OR TX "golf" OR TX "putt" OR TX "climbing" OR TX "trekking" OR TX "motor sports" OR TX "shooting" OR TX "volley ball" OR TX "orienteering" OR TX "rounders" OR TX "rowing" OR TX "triathlon" OR TX "boxing" OR TX "waterskiing" OR TX "lacrosse" OR TX "fencing" OR DE "Sport games" OR DE "Ball games" OR DE "Olympic Games" or DE "Sports" OR DE "Equestrian sport" OR DE "Football" OR DE "Baseball" OR DE "Racing" OR DE "Skis" OR DE "Martial arts" OR DE "Archery" OR DE "Fishing" OR DE "Yoga"))

OR (DE "Culture" or DE "Museums" or DE "Visual culture" or DE "Cultural behaviour" OR DE "Cultural areas" OR DE "Cultural property" OR DE "Cultural heritage" OR DE "Preservation of cultural heritage" or DE "Cultural life" OR DE "Cultural environment" OR DE "Cultural exhibitions" OR DE "Cultural expenditure" OR DE "Cultural industry" OR DE "Cultural practices" OR DE "Popular culture" or DE "Popular literature" or DE "Popular music" or DE "Popular poetry" or DE "Popular theatre" OR DE "Archaeological museums" OR DE "Art museums" OR DE "Ethnographic museums" OR DE "Galleries" OR DE "Historical museums" OR DE "Municipal museums" OR DE "Museum acquisitions" OR DE "Museum collections" OR DE "Museum objects" OR DE "National museums" OR DE "Natural history museums" OR DE "Open air museums" OR DE "Regional museums" OR DE "Science museums" OR DE "Libraries" OR DE "Map library" OR DE "Record library" OR DE "historical monuments" OR DE "historical sites" or DE "Archaeological collections" or DE "Archaeological exhibitions" or DE "Archaeological sites" OR DE "Galleries" OR DE "Arts" OR DE "Art" OR DE "Performing arts" OR DE "Visual arts" OR DE "Sculpture and carving" OR DE "Music" OR DE "Ancient music" OR DE "Choral music" OR DE "Classical music" OR DE "Contemporary music" OR DE "Dance music" OR DE "Opera" OR DE "Folk music" OR DE "Instrumental music" OR DE "Modern music" OR DE "Orchestras" OR DE "Pop music" OR DE "Reggae" OR DE "Religious music" OR DE "Ritual music" OR DE "Rock music" OR DE "Traditional music" OR DE "Vocal music" OR DE "Musical instruments" OR DE "Songs" Or DE "Dance" OR DE "Ballet" OR DE "Folk dance" OR DE "Modern dance" OR DE "Traditional dance" OR DE "Literature" OR DE "Classical literature" OR DE "Contemporary literature" OR DE "Drama" OR DE "Folk literature" OR DE "Literary works" OR DE "Novels" OR DE "Oral literature" OR DE "Poem" OR DE "Poetry" OR DE "Popular literature" OR DE "Prose" OR DE "Traditional literature" OR DE "Theatre" OR DE "Ancient theatre" OR DE "Classical theatre" OR DE "Contemporary theatre" OR DE "Mime" OR DE "National theatre" OR DE "Open air theatre" OR DE "Popular theatre" OR DE "Puppet theatre" OR DE "Shadow theatre" OR DE "Traditional theatre" OR DE "Visual arts" OR DE "Fine arts" OR DE "Graphic arts" OR DE "Iconography" OR DE "Textile arts" OR DE "Photography" OR DE "Batik" OR DE "Embroidery" OR DE "Wood-carving" OR DE "Pottery" or DE "Ceramics" OR DE "Calligraphy" OR DE "Jewellery" OR DE "Painting" OR DE "Drawing" OR DE "Carnivals" OR DE "Festivals"))

and

((DE "Engagement" or DE "Access to culture" or DE "Cultural barriers" or DE "Creativity" OR DE "Creative work" or DE "Consumption" OR DE "Cultural consumption" OR DE "Performance" OR DE "Individual performance" OR DE "Theatrical performance" OR DE "Musical performances") OR (TX participat* OR attend* OR access OR visit* OR create OR creates OR watch* OR volunteer* OR listen* OR join Or joins OR conserve OR aware* OR play OR plays OR write OR teach OR buy* OR member* OR campaign* OR coach* OR

instruct OR train* OR archive OR document OR collect OR rehearse OR make OR produce) OR (TX effect* OR impact* OR value* OR benefit* OR implication* OR advantage* OR disadvantage OR disadvantages or factor* or reason* or barrier* or facilitator*) OR (TX "take part" OR "taking part"))

Social Science Citation Index (SSCI) and Arts and Humanities Citation Index (AHCI)

TS= (engage* or participat* OR attend* OR access OR visit* OR create OR watch* OR volunteer* OR listen* OR join OR conserve OR aware* OR play OR teach OR produce OR perform)

and

(TI=(Sport or "swimming" OR "diving" OR "cycling" OR "BMX" OR "cyclo-cross" OR "biking" OR "bowls" OR "bowling" OR "keep fit" OR "aerobics" OR "dance exercise" OR "gym" OR "judo" OR "karate" OR "taekwondo" OR "martial arts" OR "self-defence" OR "tai chi" OR "weight training" OR "body building" OR "weightlifting" OR "gymnastics" OR "snooker" OR "pool" OR "billiards" OR "darts" OR "rugby" OR "football" OR "camogie" OR "hurling" OR "handball" OR "shinty" OR "cricket" OR "hockey" OR "archery" OR "baseball" OR "softball" OR "netball" OR "tennis" OR "badminton" OR "squash" OR "basketball" OR "athletics" OR "jogging" OR "cross-country" OR "running" OR "angling" OR "fishing" OR "yachting" OR "sailing" OR "canoeing" OR "windsurfing" OR "boardsailing" OR "skating" OR "curling" OR "golf" OR "pitch and putt" OR "skiing" OR "horse riding" OR "climbing" OR "mountaineering" OR "trekking" OR "motor sports" OR "shooting" OR "volleyball" OR "orienteeing" OR "rounders" OR "rowing" OR "triathlon" OR "boxing" OR "waterskiing" OR "lacrosse" OR "fencing" OR "yoga")) OR (TI= (culture or art or arts or museums or libraries or archives or galleries or heritage or (historic* city) or (historic* town) or (historic* building) or (historic* park) or (historic* garden) or (historic* landscape) or (historic* transport system) or (historic* place of worship) or (archaeological site) or (heritage site) or (historic* interest) or (historic* place) or (historic* space) or (historic* environment) or (historic* site) or (furniture making) or (jewellery making) or monument or castle* or ruin* or photography or sculpture* or video* or festival* or drama* or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or instrument or orchestra or painting* or drawing* or printmaking or film* or animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry))

British Humanities Index (BHI)

(DE=("heritage" or "culture" or "archives" or "art" or "arts" or "athletes" or "athletics" or "badminton game" or "baseball" or "basketball" or "bicycle racing" or "body building" or "bowls game" or "boxing" or "curling" or "darts" or "drama" or "fencing" or "football" or "golf" or "hockey" or "horseracing" or "horseriding" or "ice hockey" or "ice skating" or "kabadi" or "libraries" or "martial arts" or "music" or "parachuting" or "performing arts" or "polo" or "rock climbing" or "roller skating" or "rugby football" or "running" or "shooting" or "skiing" or "skydiving" or "snooker" or "sports" or "squash" or "tennis" or "theatre" or "volleyball" or "watersports" or "weightlifting" or "winter sports"))

and

(AB=("engagement" or "participate" or "visit" or "access" or "aware" or "join" or "volunteer" or "attend" or "watch" or "listen" or "perform" or "rehearse" or "play" or "write" or "create" or "purchase" or "member" or "campaign" or "teach" or "instruct" or "document" or "collect" or "produce" or "study" or "impact" or "involvement" or "go to" or "going to" or "take part" or "taking part" or "decision making" or "effect" or "value"))

Econlit

1 (engage or participate or visit or access or aware or join or volunteer or attend or watch or listen or perform or rehearse or play or write or make or create or buy or member or campaign

or teach or instruct or train or archive or document or collect or produce).mp. [mp=heading words, abstract, title, country as subject] (69996)

2 limit 1 to yr="1997 -Current" (57999)

3 (effect or value or impact).mp. [mp=heading words, abstract, title, country as subject] (128620)

4 limit 3 to yr="1997 -Current" (96143)

5 1 or 3 (180875)

6 (sport or culture or art or arts or museums or libraries or archives or galleries or heritage).mp. [mp=heading words, abstract, title, country as subject] (12439)

7 limit 6 to yr="1997 -Current" (9723)

8 (sport* or culture or art* or museum* or librar* or archive* or galler* or heritage).mp. [mp=heading words, abstract, title, country as subject] (80436)

9 limit 8 to yr="1997 -Current" (59389)

10 9 and 5 (17925)

11 7 and 5 (2781)

12 limit 11 to (yr="1997 -Current" and English) (2440)

13 from 12 keep 1-2440 (2440)

ERIC

(AB=("engagement" or "participate" or "visit" or "access" or "aware" or "join" or "volunteer" or "attend" or "watch" or "listen" or "perform" or "rehearse" or "play" or "write" or "create" or "purchase" or "member" or "campaign" or "teach" or "instruct" or "document" or "collect" or "produce" or "study" or "impact" or "involvement" or "go to" or "going to" or "take part" or "taking part" or "decision making" or "effect" or "value"))

and

(DE=("theatre arts" or "aquatic sports" or "archives" or "art" or "art history" or "athletics" or "culture" or "dance" or "drama" or "fine arts" or "folk culture" or "historic sites" or "libraries" or "museums" or "music" or "painting visual arts" or "public libraries" or "racquet sports" or "recreational activities" or "sports" or "animation"))

Medline

Set A: culture and engagement

1. *Culture/

2. *art/ or *paintings/ or *sculpture/

3. exp *libraries/ or exp *libraries, dental/ or exp *libraries, digital/ or exp *libraries, hospital/ or exp *libraries, medical/ or exp *national library of medicine (u.s.)/ or exp *libraries, nursing/

4. exp *archives/ or exp *museums/

5. engagement.mp.

6. participat*.mp.

7. ((visit* or access or aware* or volunteer* or attend* or watch* or listen* or rehears* or campaign* or 'go to' or 'tak* part') adj5 (monument or castle* or ruin* or photography or video* or festival* or drama or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or drawing* or printmaking or film* or animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry or historic* city or historic* town or historic* building or historic* park or historic* garden or historic* landscape or historic* transport system or historic* place of worship or archaeological site or heritage site or historic* interest or historic* place or historic* space or historic*

environment or historic* site or furniture making or jewellery making or galleries or gallery or heritag*)).tw.

8. ((visit* or access or aware* or volunteer* or attend* or watch* or listen* or rehears* or campaign* or 'go to' or 'tak* part') adj5 (culture or art or arts)).ti,ab.

9. 4 or 1 or 3 or 2

10. 6 or 5

11. 9 and 10

12. ((engagement or participate*) adj5 (monument or castle* or ruin* or photography or video* or festival* or drama or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or drawing* or printmaking or film* or animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry or historic* city or historic* town or historic* building or historic* park or historic* garden or historic* landscape or historic* transport system or historic* place of worship or archaeological site or heritage site or historic* interest or historic* place or historic* space or historic* environment or historic* site or furniture making or jewellery making or galleries or gallery or heritag*)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

13. ((visit* or access or aware* or volunteer* or attend* or watch* or listen* or rehears* or campaign* or 'go to' or 'tak* part') adj5 (painting or sculpture or library or libraries or archives or museums or archive or museum)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

14. 8 or 11 or 7 or 12 or 13

Set B Culture and Impact

1. *Culture/

2. *art/ or *paintings/ or *sculpture/

3. exp *libraries/ or exp *libraries, dental/ or exp *libraries, digital/ or exp *libraries, hospital/ or exp *libraries, medical/ or exp *"national library of medicine (u.s.)"/ or exp *libraries, nursing/

4. exp *archives/ or exp *museums/

5. (impact adj5 (monument or castle* or ruin* or photography or video* or festival* or drama or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or drawing* or printmaking or film* or animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

6. (benefit* adj5 (monument or castle* or ruin* or photography or video* or festival* or drama or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or drawing* or printmaking or film* or animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

7. (effect adj5 (monument or castle* or ruin* or photography or video* or festival* or drama or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or drawing* or printmaking or film* or animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

8. (value adj5 (monument or castle* or ruin* or photography or video* or festival* or drama or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or drawing* or printmaking or film* or animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

9. (economic adj5 (monument or castle* or ruin* or photography or video* or festival* or drama or theatr* or musical* or pantomime* or opera or operas or music* or jazz or ballet* or dance* or dancing or songs or singing or orchestra or drawing* or printmaking or film* or

animation* or textile* or craft* or calligraphy or pottery or story or stories or poetry)).mp.
[mp=title, original title, abstract, name of substance word, subject heading word]

10. ((impact or value or benefit or economic) adj5 (art or arts or culture)).ti.

11. ((effect or impact or value or benefit or economic) adj5 (historic* city or historic* town or historic* building or historic* park or historic* garden or historic* landscape or historic* transport system or historic* place of worship or archaeological site or heritage site or historic* interest or historic* place or historic* space or historic* environment or historic* site or furniture making or jewellery making)).tw.

12. ((effect or impact or value or benefit or economic) adj5 (galleries or gallery or heritage*)).tw.

13. (effect or impact or value or benefit or economic).ti.

14. 4 or 1 or 3 or 2

15. 14 and 13

16. 6 or 11 or 7 or 9 or 12 or 15 or 8 or 10 or 5

Set C: Sport and engagement

1. engagement.mp.

2. participat*.mp.

3. 2 or 1

4. (sport or sports).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

5. ((engagement or participat* or visit* or access or aware* or volunteer* or attend* or watch* or listen* or rehears* or campaign* or 'go to' or 'tak* part') adj5 ("cycling" or "BMX" or "cyclo-cross" or "biking" or "bowls" or "bowling" or "keep fit" or "aerobics" or "dance exercise" or "gym" or "judo" or "karate" or "taekwondo" or "taekwondo" or "tae kwon do" or "self-defence" or "weight training" or "body building" or "weightlifting" or "snooker" or "pool" or "billiards" or "darts" or "rugby" or "camogie" or "hurling" or "handball" or "shinty" or "cricket" or "archery" or "softball" or "netball" or "badminton" or "squash" or "cross-country" or "angling" or "fishing" or "yachting" or "sailing" or "canoeing" or "windsurfing" or "boardsailing" or "curling" or "pitch and putt" or "horse riding" or "climbing" or "trekking" or "motor sports" or "shooting" or "orienteering" or "rounders" or "rowing" or "triathlon" or "waterskiing" or "lacrosse" or "fencing" or "yoga" or sport or sports)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

6. (exp *dancing/ or exp *sports/ or exp *athletic performance/ or exp *baseball/ or exp *basketball/ or exp *bicycling/ or exp *boxing/ or exp *football/ or exp *golf/ or exp *gymnastics/ or exp *hockey/ or exp *martial arts/ or exp *tai ji/ or exp *mountaineering/ or exp *racquet sports/ or exp *tennis/ or exp *running/ or exp *jogging/ or exp *skating/ or exp *snow sports/ or exp *skiing/ or exp *soccer/ or exp *swimming/ or exp *diving/ or exp *"track and field"/ or exp *volleyball/ or exp *walking/ or exp *weight lifting/ or exp *wrestling/ or exp *sunbathing/ or exp survival/) not exercise*.tw.

7. 6 and 3

8. (visit* or access or aware* or volunteer* or attend* or watch* or listen* or rehears* or campaign* or 'go to' or 'tak* part').ab.ti.

9. 8 and 6

10. 7 or 9 or 5

Set D: Sport and impact

1. ((effect or impact or benefit* or effect or value or economic) adj5 ("cycling" or "BMX" or "cyclo-cross" or "biking" or "bowls" or "bowling" or "keep fit" or "aerobics" or "dance exercise" or "gym" or "judo" or "karate" or "taekwondo" or "taekwondo" or "tae kwon do" or "self-defence" or "weight training" or "body building" or "weightlifting" or "snooker" or "pool" or "billiards" or "darts" or "rugby" or "camogie" or "hurling" or "handball" or "shinty" or "cricket" or "archery" or "softball" or "netball" or "badminton" or "squash" or "cross-country" or "angling" or

"fishing" or "yachting" or "sailing" or "canoeing" or "windsurfing" or "boardsailing" or "curling" or "pitch and putt" or "horse riding" or "climbing" or "trekking" or "motor sports" or "shooting" or "orienteering" or "rounders" or "rowing" or "triathlon" or "waterskiing" or "lacrosse" or "fencing" or "yoga").mp. [mp=title, original title, abstract, name of substance word, subject heading word]

2. (effect or impact or benefit* or value or economic).ti.

3. (exp *dancing/ or exp *sports/ or exp *athletic performance/ or exp *baseball/ or exp *basketball/ or exp *bicycling/ or exp *boxing/ or exp *football/ or exp *golf/ or exp *gymnastics/ or exp *hockey/ or exp *martial arts/ or exp *tai ji/ or exp *mountaineering/ or exp *racquet sports/ or exp *tennis/ or exp *running/ or exp *jogging/ or exp *skating/ or exp *snow sports/ or exp *skiing/ or exp *soccer/ or exp *swimming/ or exp *diving/ or exp *"track and field"/ or exp *volleyball/ or exp *walking/ or exp *weight lifting/ or exp *wrestling/ or exp *sunbathing/ or exp survival/) not exercise*.tw.

4. ((effect or impact or benefit* or effect or value or economic) adj5 (sport or sports)).mp. [mp=title, original title, abstract, name of substance word, subject heading word]

5. 3 and 2

6. 5 or 1 or 4

Appendix 4: Search strings (used to identify studies from pool of 12,439)

"controlled clinical trial" or "controlled study" or "controlled trial" or "control group" or "control groups" or "experimental design" or "comparison group" or "comparison groups" or "double blind" or "placebo" or "probability sampling" or "randomised controlled trial" or "randomized controlled trial" or "random assignment" or random* or "random sampling" or "random allocation" or "single blind" or "treatment effectiveness evaluation" or "RCT" or "difference-in-difference*" or "instrumental variable*" or "propensity matching" or "case matching" or "propensity score matching" or "PSM" or "statistical adjustment" or "covariate adjustment" or "matched group*" or "statistically equated" or "cohort*" or "longitudinal" or "quasiexperiment*" or quasi experiment*" or "quasi-experiment*" or "baseline adjustment" or "pre-post" or "pre and post" or "matched variable*" or "case-mix adjustment" or "baseline comparability" or "case control" or "case-control" or "before-and-after" or "before and after" or "time series" or "time-series" or "regression discontinuity" or "nonequivalent group*" or "non-equivalent group*" or "panel stud*" or "post-hoc" or "post hoc" or "baseline" or "trial*" or "experiment*" museum* or librar* or archiv* or heritage or historic* or history or monument* or castle* or archaeolog*

Appendix 5: Definitions of learning and social impacts

1. Learning

This category includes outcomes that can be viewed as indicative of the social, affective, performative or cognitive development of an individual and/or 'proxies' for this. This includes, but is not necessarily limited to:

- academic achievement (conventional, e.g., qualifications such as GCSEs)
- transferable skills (not linked to health-related interventions/outcomes), e.g., interpersonal/communication/social competency skills
- cognitive performance (not linked to health-related interventions/outcomes)
- literacy
- numeracy
- truancy rates/behaviour problems in schools
- personal development
- cultural knowledge
- attitude to learning
- capacity or capability to learn/develop
- curiosity
- motivation for learning.

The capability to perform and/or improvements in performance at a particular sporting/cultural activity will be included as indicative of one or more of the outcomes above.

Learning associated with emotional or quality of life type impacts will be captured in the 'social impacts' category below.

2. Social

These include outcomes which can be viewed as indicative of 'well-being' (excluding health), 'community quality', and individual or community empowerment (excluding those outcomes which may be defined as knowledge/learning).

Such outcomes may include, but are not limited to:

- quality of life
- life satisfaction
- happiness
- subjective well-being
- emotional well-being
- life circumstances
- increased pride in community
- greater capacity in community
- civic well-being
- reduction in anti-social/unreasonable behaviour
- increase in socially desirable behaviour, e.g., volunteering
- self determination
- level of independence

- levels of trust
- level of reciprocal support/cooperation
- collective memory/source of intellectual ideas
- social cohesion/inclusion/solidarity
- confidence
- self-esteem
- active citizenship
- vibrancy of community
- importance of culture and sport to the community
- perception and use of built and open spaces and facilities
- community collaboration
- neighbourhood renewal
- feelgood factor
- self/community efficacy (believe that yourself or your community is more effective)
- frequency/intensity of communication between community members
- sense of place
- community cohesion
- identity (community or individual)
- well being to include: enjoyment, relaxation, escape, solace, self expression, inspiration, appreciation/understanding
- social capital
- life chances/opportunities
- national identity/pride
- democracy, decision-making
- shared experience
- community independence/reliance on government

Outcomes which can appear as the opposite of those listed above may also be included in this category, e.g., 'isolation' or 'conflict'.

Appendix 6: Map coding tool*Section A: Administration*

A.1 Name of reviewer	A.1.1 KB A.1.2 MB A.1.3 NK A.1.4 MN A.1.5 JT
A.2 Are there any linked reports?	A.2.1 Yes
A.3 Is coding based on full text?	A.3.1 Yes

Section B: Study design

<p>B.1 Study design</p> <p><i>Level 5: Random assignment of subjects to the intervention and control conditions.</i></p> <p><i>Level 4: Comparison group well-matched to intervention group on theoretically relevant factors, e.g. age, gender.</i></p> <p><i>Level 3: Comparison group present without demonstrated comparability to intervention group. Must be pre-post design.</i></p> <p><i>Level 2: One group pre-post design OR comparison group post-test only design.</i></p> <p><i>Level 1: Single-group post-test only design.</i></p>	<p>B.1.1 Systematic reviews with or without meta-analysis</p> <p>B.1.2 Level 5 Maryland Scientific Methods Scale</p> <p>B.1.3 Levels 3 or 4 Maryland Scientific Methods Scale</p> <p>B.1.4 Levels 1 or 2 Maryland Scientific Methods Scale</p>
--	--

Section C: Population

<p>C.1 What population groups are included in the study? <i>Select all relevant.</i></p>	<p>C.1.1 Children/young people (1-15 years)</p> <p>C.1.2 Adults (16+ years)</p> <p>C.1.3 Older people (65+ years)</p> <p>C.1.4 General population</p> <p>C.1.5 Unspecified</p>
<p>C.2 Does the abstract imply that there are outcome measures for any of the priority groups?</p>	<p>C.2.1 BME</p> <p>C.2.2 Women</p> <p>C.2.3 Limiting disability</p> <p>C.2.4 Low SES</p> <p>C.2.5 'Excluded groups' unspecified</p>

Section D: Engagement mode

<p>D.1 What is the engagement mode?</p>	<p>D.1.1 Attending</p> <p>D.1.2 Participating</p> <p>D.1.3 Deciding</p> <p>D.1.4 Producing</p> <p>D.1.5 Unspecified/unclear</p>
---	---

Section E: Activity/site

E.1 Culture (no further details in abstract)	E.1.1 Culture (no further details in abstract)
E.2 Sport	<p>E.2.1 American football</p> <p>E.2.2 Angling/fishing</p> <p>E.2.3 Archery</p> <p>E.2.4 Badminton</p> <p>E.2.5 Basketball</p> <p>E.2.6 Baseball/Softball</p> <p>E.2.7 BMX/Cyclo-cross/mountain biking</p> <p>E.2.8 Bowls (<i>indoor and lawn, not tenpin</i>)</p> <p>E.2.9 Boxing</p> <p>E.2.10 Canoeing</p> <p>E.2.11 Climbing/mountaineering</p> <p>E.2.12 Cricket</p> <p>E.2.13 Curling</p> <p>E.2.14 Cycling (<i>for health, recreation, training competition</i>)</p> <p>E.2.15 Darts</p> <p>E.2.16 Football (<i>soccer</i>)</p> <p>E.2.17 Fencing</p> <p>E.2.18 Gaelic sports (<i>includes: camogie, shinty, hurling Gaelic football</i>)</p> <p>E.2.19 Golf/pitch and putt/putting</p> <p>E.2.20 Gym or conditioning activities</p> <p>E.2.21 Gymnastics</p> <p>E.2.22 Hill trekking or backpacking</p> <p>E.2.23 Hockey</p> <p>E.2.24 Horse riding (<i>i.e., equestrian activities. Equestrianism refers to the skill of riding or driving horses. Includes polo.</i>)</p> <p>E.2.25 Ice skating</p> <p>E.2.26 Jogging, cross country, road running</p>

	E.2.27 Judo
	E.2.28 Karate
	E.2.29 Keep fit/aerobics/dance exercise <i>(includes exercise bike)</i>
	E.2.30 Lacrosse
	E.2.31 Martial arts - other <i>(includes self defence and tai chi)</i>
	E.2.32 Motor sports
	E.2.33 Netball
	E.2.34 Orienteering
	E.2.35 Rounders
	E.2.36 Rowing
	E.2.37 Rugby <i>(union or league)</i>
	E.2.38 Sailing/yachting
	E.2.39 Shooting <i>(air, clay target, crossbow, muzzle loading, pistol, rifle and target)</i>
	E.2.40 Skiing
	E.2.41 Snooker/billiards/pool
	E.2.42 Squash
	E.2.43 Swimming or diving <i>(indoors or outdoors)</i>
	E.2.44 Table tennis
	E.2.45 Taekwondo
	E.2.46 Tenpin bowling
	E.2.47 Tennis
	E.2.48 Track and field athletics
	E.2.49 Triathlon
	E.2.50 Volleyball
	E.2.51 Waterskiing
	E.2.52 Weightlifting
	E.2.53 Weight training <i>(includes body</i>

	<p><i>building)</i></p> <p>E.2.54 Windsurfing or boardsailing</p> <p>E.2.55 Yoga</p> <p>E.2.56 Sport (unspecified)</p>
E.3 Heritage/historic environment	E.3.1 Heritage (<i>please specify</i>)
E.4 MLA	<p>E.4.1 Museums</p> <p>E.4.2 Libraries</p> <p>E.4.3 Archives</p>
E.5 Arts attendance	<p>E.5.1 Exhibition of collection of art, photography or sculpture</p> <p>E.5.2 Event which includes video art or electronic art</p> <p>E.5.3 Culturally specific festival (e.g. mela, baisakhi, navratri)</p> <p>E.5.4 Play/drama</p> <p>E.5.5 Other/unspecified theatre performance (e.g. musical, pantomime)</p> <p>E.5.6 Opera/operetta</p> <p>E.5.7 Classical music performance</p> <p>E.5.8 Jazz performance</p> <p>E.5.9 Other/unspecified live music event</p> <p>E.5.10 Ballet</p> <p>E.5.11 Contemporary dance</p> <p>E.5.12 African people's dance or South Asian and Chinese dance</p> <p>E.5.13 Other/unspecified live dance event</p> <p>E.5.14 Arts (unspecified)</p>
E.6 Arts participation	<p>E.6.1 Ballet</p> <p>E.6.2 Other dance (not for fitness)</p> <p>E.6.3 Sing to an audience or rehearse for a performance (not karaoke)</p>

	<p>E.6.4 Play a musical instrument to an audience or rehearse for a performance</p> <p>E.6.5 Play a musical instrument for pleasure</p> <p>E.6.6 Write music</p> <p>E.6.7 Rehearse or perform play/drama</p> <p>E.6.8 Rehearse or perform opera/opera/operetta</p> <p>E.6.9 Paint, draw, printmaking or sculpture</p> <p>E.6.10 Photography as an artistic activity (not family or holiday 'snaps')</p> <p>E.6.11 Make films or videos as an artistic activity (not family or holiday)</p> <p>E.6.12 Use computer to create original artworks or animation</p> <p>E.6.13 Textile crafts such as embroidery, crocheting or knitting</p> <p>E.6.14 Wood crafts such as wood turning, carving or furniture making</p> <p>E.6.15 Other crafts such as calligraphy, pottery or jewellery making</p> <p>E.6.16 Write any stories or plays</p> <p>E.6.17 Write poetry</p> <p>E.6.18 Arts (unspecified)</p>
--	---

Section F: Outcomes

<p>F.1 What outcomes have been measured?</p>	<p>F.1.1 Learning/skills/training</p> <p>F.1.2 Social</p> <p>F.1.3 Economic (inc. employment)</p> <p>F.1.4 "Willingness to pay"</p> <p>F.1.5 Environmental</p> <p>F.1.6 Other (please specify)</p> <p>F.1.7 Unclear/not stated</p>
--	--

Appendix 7: Studies for which reports could not be obtained

Coronado AA (2000) The effects of a summer performing arts program on at-risk adolescents. California: *ProQuest Information & Learning, Dissertation Abstracts International Section A: Humanities and Social Sciences* M1-60, 4358-4358.

Jergovic D (2001) The impact of athletic participation on the academic achievement of American adolescents. Chicago: *ProQuest Information & Learning, Dissertation Abstracts International: Section B: The Sciences and Engineering* M1-62, 576-576.

McClendon CY (2000) Promoting achievement in school through sports (PASS): An evaluation study. Maryland: *ProQuest Information & Learning, Dissertation Abstracts International Section A: Humanities and Social Sciences* M1-60, 2372-2372.

Morand MK (2004) The effects of mixed martial arts and exercise on behavior of boys with attention deficit hyperactivity disorder. Hempstead, NY: *ProQuest Information & Learning, Dissertation Abstracts International: Section B: The Sciences and Engineering* M1-65, 2609-2609.

Poulsen JCS (1999) Efficacy of drama based teaching on children with learning disabilities. Calgary: *ProQuest Information & Learning, Dissertation Abstracts International Section A: Humanities and Social Sciences* M1-59, 4360-4360.

Seham JC (1998) The effects on at-risk children of an in-school dance program. New York: *ProQuest Information & Learning, Dissertation Abstracts International: Section B: The Sciences and Engineering* M1-58, 4471-4471.

Appendix 8: Identifying further studies through checking of relevant systematic reviews

Citation checking of the following reviews was undertaken to identify further relevant literature for the in-depth reviews.

Sport and arts

* indicates the original three (arts participation) reviews from which the other reviews listed here were 'snowballed'.

Butzlaff R (2000) Can music be used to teach reading? *Journal of Aesthetic Education* 34(3-4):167-178.

*Gold C, Voracek M, Wigram T (2004) Effects of music therapy for children and adolescents with psychopathology: a meta-analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 45(6): 1054-1063.

Gold C, Wigram T, Elefant C (2006) *Music therapy for autistic spectrum disorder*. Cochrane Database of Systematic Reviews 2006, Issue 2. Art. No.: CD004381. DOI: 10.1002/14651858.CD004381.pub2.

*Hetland L, Winner E (2001) The arts and academic achievement: what the evidence shows. *Arts Education Policy Review* 102(5): 3-6.

Keinanen M, Hetland L, Winner E (2000) Teaching cognitive skill through dance: evidence for near but not far transfer. *Journal of Aesthetic Education* 34(3-4): 295-306.

Moga E, Burger K, Hetland L, Winner E (2000) Does studying the arts engender creative thinking? Evidence for near but not far transfer. *Journal of Aesthetic Education* 34(3-4): 91-104.

Podlozny A (2000) Strengthening verbal skills through the use of classroom drama: a clear link. *Journal of Aesthetic Education* 34(3-4): 239-275.

*Standley J (2008) Does music instruction help children learn to read? Evidence of a meta-analysis. *Update: Applications of Research in Music Education* 27(1): 17-32.

Vaughn K (2000) Music and mathematics: modest support for the oft-claimed relationship. *Journal of Aesthetic Education* 34(3-4): 149-166.

*Winner E, Cooper M (2000) Mute those claims: no evidence (yet) for a causal link between arts study and academic achievement. *Journal of Aesthetic Education* 34(3-4): 11-75.

Museums and libraries

Ecsite-UK (2008) *The impact of science and discovery centres: a review of worldwide studies*. Bristol: European Collaboration of Science, Industry and Technology Exhibitions.

Wavell C, Baxter G, Johnson I, Williams D (2002) *Impact evaluation of museums, archives and libraries: available evidence project*. London: Resource: The Council for Museums, Archives and Libraries.

Williams D, Coles L, Wavell C (2002) *Impact of school library services on achievement and learning in primary schools: critical literature review of the impact of school library provision on achievement and learning in primary level students*. London: Museums Libraries and Archives Council.

Williams D, Wavell C, Coles L (2001) *Impact of school library services on achievement and learning*. London: Museums Libraries and Archives Council.

Appendix 9: In-depth review data extraction and quality assessment tool

Section A: Administrative details

A.1 Name of the reviewer	A.1.1 KB A.1.2 MN A.1.3 JT A.1.4 CV A.1.5 IK A.1.6 NK
A.2 Date of the review	A.2.1 Details
A.3 Are there any known linked reports?	A.3.1 No A.3.2 Yes (please specify)

Section B: Study details

B.1 What are the broad aims of the study?	B.1.1 Explicitly stated (please specify) B.1.2 Implicit (please specify) B.1.3 Unclear/not stated (please specify)
B.2 In which country/countries was the study conducted?	B.2.1 UK (please specify) B.2.2 Europe (please specify) B.2.3 Scandinavia (please specify) B.2.4 Russia B.2.5 USA B.2.6 Canada B.2.7 Australia/New Zealand (please specify) B.2.8 Middle East B.2.9 Asia (please specify) B.2.10 Africa (please specify) B.2.11 Central/South America (please specify)

	B.2.12 Other (please specify)
--	-------------------------------

Section C: Participants

C.1 Who participated in the study?	<p>C.1.1 Children/young people</p> <p>C.1.2 Adults</p> <p>C.1.3 Older people (65+)</p> <p>C.1.4 Unclear (please specify)</p> <p>C.1.5 Not stated</p>
C.2 Number of participants	<p>C.2.1 Details</p> <p>C.2.2 Unclear (please specify)</p> <p>C.2.3 Not stated</p>
<p>C.3 Age of participants</p> <p><i>If the ages of participants are not stated but the authors refer to year group then report these details.</i></p>	<p>C.3.1 0-5 years</p> <p>C.3.2 6-10 years</p> <p>C.3.3 11-15 years</p> <p>C.3.4 16-18 years</p> <p>C.3.5 19+ years</p> <p>C.3.6 Unclear (please specify)</p> <p>C.3.7 Not stated</p>
C.4 Type of educational institution attended	<p>C.4.1 Pre-school/nursery/kindergarten (please specify)</p> <p>C.4.2 Primary school (please specify)</p> <p>C.4.3 Secondary school (please specify)</p> <p>C.4.4 Post-16 education (sixth-form, FE college) (please specify)</p> <p>C.4.5 Unclear (please specify)</p> <p>C.4.6 Not stated</p>
C.5 Sex of participants	<p>C.5.1 Female</p> <p>C.5.2 Male</p> <p>C.5.3 Mixed sex</p>

	C.5.4 Not stated
<p>C.6 Please specify any other important features of the participants <i>Please provide any further relevant details about the people in the sample; for example, if they have a disability.</i></p> <p><i>Also note if the participants in the intervention/control groups are split over several sites (e.g. more than one school, or in different cities).</i></p>	C.6.1 Details

Section D: Engagement mode

<p>D.1 What is the engagement mode? <i>If the intervention incorporates both attendance and participation, then select the third item ('Both attendance and participation').</i></p>	<p>D.1.1 Attending</p> <p>D.1.2 Participating</p> <p>D.1.3 Both attending and participating (in same intervention)</p> <p>D.1.4 Unspecified</p>
--	---

Section E: Sector

<p>E.1 Which sectors does the engagement relate to?</p>	<p>E.1.1 Sport</p> <p>E.1.2 Arts</p> <p>E.1.3 Museums</p> <p>E.1.4 Libraries</p> <p>E.1.5 Archives</p> <p>E.1.6 Heritage</p> <p>E.1.7 Culture (unspecified)</p>
---	---

Section F: Activity/intervention

<p>F.1 Sport <i>If the intervention is multi-component, then select F.1.58 and write each of the activities in the text box. For example, an after-school sports programme might involve a range of different sports, rather than being focused on just one sport.</i></p>	<p>F.1.1 American football</p> <p>F.1.2 Angling/fishing</p> <p>F.1.3 Archery</p> <p>F.1.4 Badminton</p> <p>F.1.5 Basketball</p> <p>F.1.6 Baseball/Softball</p> <p>F.1.7 BMX/Cyclo-cross/mountain biking</p> <p>F.1.8 Bowls (<i>indoor and lawn, not tenpin</i>)</p> <p>F.1.9 Boxing</p> <p>F.1.10 Canoeing</p> <p>F.1.11 Climbing/mountaineering</p> <p>F.1.12 Cricket</p> <p>F.1.13 Curling</p> <p>F.1.14 Cycling (<i>for health, recreation,</i></p>
--	--

	<p><i>training competition)</i></p> <p>F.1.15 Darts</p> <p>F.1.16 Football (<i>soccer</i>)</p> <p>F.1.17 Fencing</p> <p>F.1.18 Gaelic sports (<i>includes: camogie, shinty, hurling Gaelic football</i>)</p> <p>F.1.19 Golf/pitch and putt/putting</p> <p>F.1.20 Gym or conditioning activities</p> <p>F.1.21 Gymnastics</p> <p>F.1.22 Hill trekking or backpacking</p> <p>F.1.23 Hockey</p> <p>F.1.24 Horse riding (<i>i.e., equestrian activities. Equestrianism refers to the skill of riding or driving horses. Includes polo.</i>)</p> <p>F.1.25 Ice skating</p> <p>F.1.26 Jogging, cross country, road running</p> <p>F.1.27 Judo</p> <p>F.1.28 Karate</p> <p>F.1.29 Keep fit/aerobics/dance exercise (<i>includes exercise bike</i>)</p> <p>F.1.30 Lacrosse</p> <p>F.1.31 Martial arts - other (<i>includes self defence and tai chi</i>)</p> <p>F.1.32 Motor sports</p> <p>F.1.33 Netball</p> <p>F.1.34 Orienteering</p> <p>F.1.35 Rounders</p> <p>F.1.36 Rowing</p> <p>F.1.37 Rugby (<i>union or league</i>)</p> <p>F.1.38 Sailing/yachting</p> <p>F.1.39 Shooting (<i>air, clay target, crossbow, muzzle loading, pistol, rifle and target</i>)</p>
--	---

	<p>F.1.40 Skiing</p> <p>F.1.41 Snooker/billiards/pool</p> <p>F.1.42 Squash</p> <p>F.1.43 Swimming or diving (<i>indoors or outdoors</i>)</p> <p>F.1.44 Table tennis</p> <p>F.1.45 Taekwondo</p> <p>F.1.46 Tenpin bowling</p> <p>F.1.47 Tennis</p> <p>F.1.48 Track and field athletics</p> <p>F.1.49 Triathlon</p> <p>F.1.50 Volleyball</p> <p>F.1.51 Waterskiing</p> <p>F.1.52 Weightlifting</p> <p>F.1.53 Weight training (<i>includes body building</i>)</p> <p>F.1.54 Windsurfing or boardsailing</p> <p>F.1.55 Yoga</p> <p>F.1.56 Sport (other) (please specify) <i>NB: as this sport is not on the Taking Part list, the results are not relevant for this review.</i></p> <p>F.1.57 Sport (unspecified) (please specify)</p> <p>F.1.58 Sport (multi-component) (please specify) <i>Only select this if the intervention involves more than one sport. If the intervention involves sport(s) and another component, such as life skills, this will be picked up in section G.</i></p>
<p>F.2 Arts attendance</p> <p><i>If the intervention is multi-component, then select F.2.15 and write each of the activities in the text box. For example, a performing arts project might involve two different activities: (1) attend play (2) perform dance</i></p>	<p>F.2.1 Exhibition of collection of art, photography or sculpture</p> <p>F.2.2 Event which includes video art or electronic art</p> <p>F.2.3 Culturally specific festival (e.g. mela, baisakhi, navratri)</p>

	<p>F.2.4 Play/drama</p> <p>F.2.5 Other/unspecified theatre performance (e.g. musical, pantomime)</p> <p>F.2.6 Opera/opera</p> <p>F.2.7 Classical music performance</p> <p>F.2.8 Jazz performance</p> <p>F.2.9 Other/unspecified live music event</p> <p>F.2.10 Ballet</p> <p>F.2.11 Contemporary dance</p> <p>F.2.12 African people's dance or South Asian and Chinese dance</p> <p>F.2.13 Other/unspecified live dance event</p> <p>F.2.14 Arts (unspecified)</p> <p>F.2.15 Arts (multi-component) (please specify) <i>Only select this if the intervention involves more than one arts. If the intervention involves art(s) and another component, such as life skills, this will be picked up in section G.</i></p>
<p>F.3 Arts participation <i>If the intervention is multi-component, then select F.3.19 and write each of the activities in the text box. For example, a performing arts project might involve four different activities:</i></p> <ol style="list-style-type: none"> 1. <i>writing plays</i> 2. <i>rehearsing/performing plays</i> 3. <i>singing</i> 4. <i>dancing.</i> 	<p>F.3.1 Ballet</p> <p>F.3.2 Other dance (not for fitness)</p> <p>F.3.3 Sing to an audience or rehearse for a performance (not karaoke)</p> <p>F.3.4 Play a musical instrument to an audience or rehearse for a performance</p> <p>F.3.5 Play a musical instrument for pleasure</p> <p>F.3.6 Write music</p> <p>F.3.7 Rehearse or perform play/drama</p> <p>F.3.8 Rehearse or perform opera/opera</p> <p>F.3.9 Paint, draw, printmaking or sculpture</p> <p>F.3.10 Photography as an artistic activity (not family or holiday 'snaps')</p> <p>F.3.11 Make films or videos as an artistic activity (not family or holiday)</p>

	<p>F.3.12 Use computer to create original artworks or animation</p> <p>F.3.13 Textile crafts such as embroidery, crocheting or knitting</p> <p>F.3.14 Wood crafts such as wood turning, carving or furniture making</p> <p>F.3.15 Other crafts such as calligraphy, pottery or jewellery making</p> <p>F.3.16 Write any stories or plays</p> <p>F.3.17 Write poetry</p> <p>F.3.18 Arts (unspecified)</p> <p>F.3.19 Arts (multi-component) (please specify) <i>Only select this if the intervention involves more than one arts. If the intervention involves art(s) and another component, such as life skills, this will be picked up in section G.</i></p>
--	---

Section G: Activity/intervention description

<p>G.1 Does the activity/intervention have a formal name?</p>	<p>G.1.1 Yes (please specify)</p> <p>G.1.2 No</p> <p>G.1.3 Unclear (please specify)</p>
<p>G.2 If it has a formal name, at what level is the activity/intervention offered? (i.e., who is it available to?) <i>The first items all refer to geographical regions.</i> 1. <i>national</i> 2. <i>regional (e.g., Tayside, Yorkshire, California)</i> 3. <i>local (e.g., Edinburgh)</i> 4. <i>community (e.g., Bangladeshi community in Tower Hamlets)</i></p> <p><i>Organisational/institutional should be selected when the intervention is not available to everyone in an area, rather it is only offered at the level of a particular institution or organisation, such as a single school.</i></p>	<p>G.2.1 Not applicable (not a named strategy/programme etc.)</p> <p>G.2.2 National</p> <p>G.2.3 Regional</p> <p>G.2.4 Local</p> <p>G.2.5 Community</p> <p>G.2.6 Organisational/institutional</p> <p>G.2.7 Structural</p> <p>G.2.8 Unclear (please specify)</p> <p>G.2.9 Not stated</p>
<p>G.3 Aim(s) of the activity/intervention</p>	<p>G.3.1 Explicitly stated (please specify)</p> <p>G.3.2 Implicit (please specify)</p> <p>G.3.3 Not stated</p>
<p>G.4 Is the activity/intervention multi-component or involve a single activity? <i>For example, a multi-component arts therapy programme might involve drawing, painting and singing.</i></p>	<p>G.4.1 Single (please provide very brief description)</p> <p>G.4.2 Multi-component (please provide very brief description)</p> <p>G.4.3 Unclear (please specify)</p> <p>G.4.4 Not stated</p>
<p>G.5 If a multi-component intervention, does it involve a non-sport/arts/MLA/heritage component (e.g., life skills)?</p>	<p>G.5.1 Not applicable (single activity)</p> <p>G.5.2 Yes, intervention has a non-arts/sports/MLA/heritage component (please specify)</p> <p>G.5.3 No, all components are arts/sports/MLA/heritage</p> <p>G.5.4 Unclear (please specify)</p> <p>G.5.5 Not stated</p>

<p>G.6 Is the activity/intervention targeted at populations with particular characteristics?</p>	<p>G.6.1 Academic underachievement</p> <p>G.6.2 Limited English proficiency/English as additional language</p> <p>G.6.3 Low SES</p> <p>G.6.4 Limiting disability</p> <p>G.6.5 Other (please specify)</p> <p>G.6.6 Unclear (please specify)</p> <p>G.6.7 No</p> <p>G.6.8 Not stated</p>
<p>G.7 Do participants receive the activity/intervention individually or as a group?</p>	<p>G.7.1 Individual</p> <p>G.7.2 Group</p> <p>G.7.3 Unclear (please specify)</p> <p>G.7.4 Not stated</p>
<p>G.8 What is/are the setting(s) of the activity/intervention?</p>	<p>G.8.1 Home</p> <p>G.8.2 Community centre</p> <p>G.8.3 Sports centre</p> <p>G.8.4 Other sport setting (please specify)</p> <p>G.8.5 Church-based setting</p> <p>G.8.6 Club (context unspecified): e.g., "youth club" (please specify)</p> <p>G.8.7 Further education institution (e.g., college)</p> <p>G.8.8 School (in school hours)</p> <p>G.8.9 School-based extra-curricular clubs (i.e. those taking place in schools in out-of-school hours)</p> <p>G.8.10 Other educational setting (please specify)</p> <p>G.8.11 Hospital</p> <p>G.8.12 Other health setting (please specify)</p> <p>G.8.13 Music conservatory</p>

	<p>G.8.14 Museum (art) (please specify)</p> <p>G.8.15 Museum (science) (please specify)</p> <p>G.8.16 Museum (natural history) (please specify)</p> <p>G.8.17 Museum (historical) (please specify)</p> <p>G.8.18 Museum (other/unspecified) (please specify)</p> <p>G.8.19 Library (school/college) (please specify)</p> <p>G.8.20 Library (public) (please specify)</p> <p>G.8.21 Library (other/unspecified) (please specify)</p> <p>G.8.22 Archive (please specify)</p> <p>G.8.23 Heritage site (please specify)</p> <p>G.8.24 Arts setting (e.g. theatre/gallery) (please specify)</p> <p>G.8.25 Other (please specify)</p> <p>G.8.26 Unclear (please specify)</p> <p>G.8.27 Not stated</p>
<p>G.9 Who delivered the activity/intervention? <i>This question refers to the persons who actually met with the children/young people and facilitated their involvement with the activities (i.e., it doesn't mean the people/organisations who designed/developed/funded the intervention – although these may actually be the same people). Tick as many as appropriate.</i></p>	<p>G.9.1 Health professional (please specify)</p> <p>G.9.2 Parent</p> <p>G.9.3 Peer</p> <p>G.9.4 Researcher</p> <p>G.9.5 Social worker</p> <p>G.9.6 Youth worker</p> <p>G.9.7 Teacher/tutor/lecturer</p> <p>G.9.8 Other educational specialist (e.g., education officer at a museum) - please specify</p> <p>G.9.9 Librarian</p> <p>G.9.10 Curator</p> <p>G.9.11 Artist (broadly defined) (please specify)</p>

	<p>G.9.12 Sports professional (please specify)</p> <p>G.9.13 Other (please specify)</p> <p>G.9.14 Unclear (please specify)</p> <p>G.9.15 Not stated</p>
<p>G.10 Is the activity/intervention led by a trained or untrained person?</p>	<p>G.10.1 Trained</p> <p>G.10.2 Untrained</p> <p>G.10.3 Neither (e.g. guidance device) (please specify)</p> <p>G.10.4 Unclear (please specify)</p> <p>G.10.5 Not stated</p>
<p>G.11 Are those delivering the activity/intervention doing so on a volunteer basis, or are they employed?</p>	<p>G.11.1 Employed</p> <p>G.11.2 Volunteer</p> <p>G.11.3 Unclear (please specify)</p> <p>G.11.4 Not stated</p>
<p>G.12 Duration of the activity/intervention <i>Choose the relevant category and write in the exact intervention length if specified in the report</i></p> <p><i>When the intervention is ongoing, tick 'OTHER' and indicate the length of intervention as the length of the outcome assessment period</i></p>	<p>G.12.1 One day or less (please specify)</p> <p>G.12.2 1 day to 1 week (please specify)</p> <p>G.12.3 1 week (and 1 day) to 1 month (please specify)</p> <p>G.12.4 1 month (and 1 day) to 3 months (please specify)</p> <p>G.12.5 3 months (and 1 day) to 6 months (please specify)</p> <p>G.12.6 6 months (and 1 day) to 1 year (please specify)</p> <p>G.12.7 1 year (and 1 day) to 2 years (please specify)</p> <p>G.12.8 2 years (and 1 day) to 3 years (please specify)</p> <p>G.12.9 3 years (and 1 day) to 5 years (please specify)</p> <p>G.12.10 more than 5 years (please specify)</p> <p>G.12.11 Other (please specify)</p>

	<p>G.12.12 Unclear (please specify)</p> <p>G.12.13 Not stated</p>
G.13 Intensity of the activity/intervention	<p>G.13.1 Once</p> <p>G.13.2 Daily</p> <p>G.13.3 1-2 per week</p> <p>G.13.4 3-4 per week</p> <p>G.13.5 5-7 days per week (please specify)</p> <p>G.13.6 2-3 times per month (please specify)</p> <p>G.13.7 Monthly</p> <p>G.13.8 6-monthly (i.e. twice a year)</p> <p>G.13.9 Annually</p> <p>G.13.10 Other (please specify)</p> <p>G.13.11 Unclear (please specify)</p> <p>G.13.12 Not stated</p>
G.14 Please specify any other relevant information about the activity/intervention	G.14.1 Details

Section H: Study design/methods

<p>H.1 What was the design of the evaluation? <i>To what degree have the studies controlled for bias (through the study design used) that might invalidate the results?</i></p>	<p>H.1.1 Randomised controlled trial (Maryland Scale 5) <i>Not necessarily a pre-post design.</i></p> <p>H.1.2 Well-matched comparison group study (Maryland Scale 4) <i>Comparison group well matched to intervention group on theoretically relevant factors (e.g. age, gender, etc.). Post-hoc statistical analysis used to control for differences (i.e. to match the groups). For example, study may have used Propensity Score Matching.</i></p> <p><i>Studies classed as Maryland Scale 4 design must be pre-post.</i></p> <p>H.1.3 Unmatched comparison group study (Maryland Scale 3) <i>Comparison group present without demonstrated comparability to intervention group. Must be pre-post design.</i></p> <p>H.1.4 One group pre-post study OR comparison group post-test only (Maryland Scale 2) <i>NB: some studies might be longitudinal but there is only one measurement (post-test) of the outcome of interest (e.g. studies using modelling and retrospective data and measuring effect of teenage sport participation on adult outcomes such as Bachelor degrees).</i></p> <p>H.1.5 Single-group post-test study (Maryland Scale 1)</p>
<p>H.2 What treatment/intervention did the control/comparison group receive?</p>	<p>H.2.1 Not applicable (one group only)</p> <p>H.2.2 No treatment</p> <p>H.2.3 Treatment as usual (e.g. normal PE lessons) (please specify)</p> <p>H.2.4 Alternative intervention (please specify)</p> <p>H.2.5 Unclear (please specify)</p> <p>H.2.6 Not stated</p>
<p>H.3 What was the unit of allocation to the intervention/control groups?</p>	<p>H.3.1 Not applicable (one group only)</p> <p>H.3.2 Individual students</p>

	<p>H.3.3 Classes</p> <p>H.3.4 Schools</p> <p>H.3.5 Other (please specify)</p> <p>H.3.6 Unclear (please specify)</p> <p>H.3.7 Not stated</p>
H.4 Which methods were used to collect the data?	<p>H.4.1 Questionnaire/survey instrument completed by student</p> <p>H.4.2 Questionnaire/survey instrument completed by teacher, by parent, etc. (please specify)</p> <p>H.4.3 Self-completion report or diary</p> <p>H.4.4 One-to-one interview (face-to-face, telephone)</p> <p>H.4.5 Focus group interview</p> <p>H.4.6 Observation</p> <p>H.4.7 Curriculum/subject-based test (assessment/exam etc. completed by student)</p> <p>H.4.8 Clinical test</p> <p>H.4.9 Practical test</p> <p>H.4.10 Psychological test (e.g. IQ test)</p> <p>H.4.11 Hypothetical scenario, including vignettes</p> <p>H.4.12 School/college records (e.g. attendance, examination results)</p> <p>H.4.13 Secondary datasets (e.g. National Educational Longitudinal Study)</p> <p>H.4.14 Other (please specify)</p> <p>H.4.15 Unclear (please specify)</p> <p>H.4.16 Not stated</p> <p>H.4.17 Please specify any other important features of data collection</p>
H.5 Which methods were used to analyse the data?	<p>H.5.1 Explicitly stated (please specify)</p>

	H.5.2 Implicit (please specify) H.5.3 Unclear (please specify) H.5.4 Not stated
H.6 Please specify any other relevant information about the design/methods	H.6.1 Details

Section I: Outcomes

<p>I.1 What learning outcomes does the study measure/report?</p>	<p>I.1.1 Academic achievement/skills (e.g., qualifications such as GCSEs or class test scores) (please specify)</p> <p>I.1.2 Transferable skills (e.g., interpersonal/communication skills, social competency/skills) (please specify)</p> <p>I.1.3 Cognitive performance (not linked to health-related intervention/outcome) (please specify)</p> <p>I.1.4 Truancy rates/behaviour problems (please specify)</p> <p>I.1.5 Personal development (please specify)</p> <p>I.1.6 Cultural knowledge (please specify)</p> <p>I.1.7 Attitude to learning (please specify)</p> <p>I.1.8 Capacity or capability to learn/develop (please specify)</p> <p>I.1.9 Curiosity (please specify)</p> <p>I.1.10 Motivation for learning (please specify)</p> <p>I.1.11 Creativity (please specify)</p> <p>I.1.12 Other (e.g., attainment expectations) (please specify)</p> <p>I.1.13 Unclear (please specify)</p> <p>I.1.14 Not stated</p>
<p>I.2 If academic achievement is measured, what subjects does the study focus on?</p>	<p>I.2.1 Not applicable (academic achievement not measured)</p> <p>I.2.2 Literacy</p> <p>I.2.3 Numeracy</p> <p>I.2.4 Other (please specify)</p> <p>I.2.5 Unclear (please specify)</p>

	I.2.6 Not stated
I.3 What non-learning outcomes does the study measure/report?	<p>I.3.1 Engagement (i.e., attendance or participation in sports/arts/MLA/heritage activities/sites) (please specify) <i>In results section, make note of whether the intervention worked to drive up engagement or not.</i></p> <p>I.3.2 Health (please specify)</p> <p>I.3.3 Other (please specify)</p> <p>I.3.4 None</p>

Section J: Quality assessment

<p>J.1 Was the allocation sequence adequately generated? Was selection bias adequately addressed?</p>	<p>J.1.1 Yes <i>Criteria for a judgement of 'YES' (i.e., low risk of bias).</i></p> <p><i>The investigators describe a random component in the sequence generation process such as:</i></p> <p><i>Referring to a random number table;</i> <i>Using a computer random number generator;</i> <i>Coin tossing;</i> <i>Shuffling cards or envelopes;</i> <i>Throwing dice;</i> <i>Drawing of lots;</i> <i>Minimization*.</i></p> <p><i>*Minimization may be implemented without a random element, and this is considered to be equivalent to being random.</i></p> <p>J.1.2 No <i>Criteria for the judgement of 'NO' (i.e., high risk of bias).</i></p> <p><i>The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example:</i></p> <ul style="list-style-type: none"> • Sequence generated by odd or even date of birth; • Sequence generated by some rule based on date (or day) of admission; • Sequence generated by some rule based on hospital or clinic record number. <p><i>Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgement or some method of non-random categorization of participants, for example:</i></p> <ul style="list-style-type: none"> • Allocation by judgement of the clinician; • Allocation by preference of the participant; • Allocation based on the results of a laboratory test or a series of tests; • Allocation by availability of the intervention. <p>J.1.3 Unclear <i>Criteria for the judgement of 'UNCLEAR'</i></p>
---	---

	<p><i>(uncertain risk of bias).</i></p> <p><i>Insufficient information about the sequence generation process to permit judgement of 'Yes' or 'No'.</i></p> <p>J.1.4 Not applicable (Study not coded as MSMS 4 or 5)</p>
J.2 Was allocation adequately concealed?	<p>J.2.1 Yes <i>Criteria for a judgement of 'YES' (i.e., low risk of bias).</i></p> <p><i>Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:</i></p> <ul style="list-style-type: none"> • Central allocation (including telephone, web-based and pharmacy-controlled randomisation); • Sequentially numbered drug containers of identical appearance; • Sequentially numbered, opaque, sealed envelopes. <p>J.2.2 No <i>Criteria for the judgement of 'NO' (i.e., high risk of bias).</i></p> <p><i>Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on:</i></p> <ul style="list-style-type: none"> • Using an open random allocation schedule (e.g. a list of random numbers); • Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered); • Alternation or rotation; • Date of birth; • Case record number; • Any other explicitly unconcealed procedure. <p>J.2.3 Unclear <i>Criteria for the judgement of 'UNCLEAR' (uncertain risk of bias).</i></p> <p><i>Insufficient information to permit judgement of 'Yes' or 'No'. This is usually the case if</i></p>

	<p><i>the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.</i></p> <p>J.2.4 Not applicable (Study not coded as MSMS 4 or 5)</p>
<p>J.3 Was knowledge of the allocation to groups adequately prevented? (Blinding) <i>Was knowledge of the allocation to intervention and control groups adequately prevented during the study (i.e. were relevant people blinded to this information)?</i></p> <p><i>Reviewers should consider blinding of participants, of personnel such as those delivering the intervention, and of those assessing the outcomes.</i></p> <p><i>What measures were used, if any, to blind study participants/personnel/assessors from knowledge of which intervention a participant received? Is there any information relating to whether the intended blinding was effective?</i></p>	<p>J.3.1 Not relevant (only one group) (2)</p> <p>J.3.2 Yes (1) <i>Use this code for any of the following:</i></p> <ul style="list-style-type: none"> • No blinding, but reviewers judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding. • Blinding of participants and key personnel ensured, and unlikely that the blinding could have been broken; • Either participants or some key study personnel were not blinded, but outcome measurement was blinded and the non-blinding of others unlikely to introduce bias. <p>NB: Blinding is not relevant where outcome is based on official statistics/centrally administered tests (e.g. GCSE results); nor is blinding relevant where study used a single group design; e.g. outcome measured using a cross sectional survey (because outcome measure cannot be manipulated and/or no incentive to manipulate it).</p> <p>J.3.3 No (2) <i>Any one of the following:</i></p> <ul style="list-style-type: none"> • No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding; • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken; • Either participants or some key study personnel were not blinded, and the non-blinding of others likely to introduce bias. <p>J.3.4 Unclear (2) <i>Insufficient information to permit judgement of 'Yes' or 'No'.</i></p>

<p>J.4 Was incomplete outcome data addressed? <i>This question is concerned with the completeness of outcome data and whether the issue of any incomplete outcome data has been adequately addressed.</i></p>	<p>J.4.1 Yes (1) <i>Criteria for a judgement of 'YES' (i.e., low risk of bias).</i></p> <p><i>Any one of the following:</i></p> <ul style="list-style-type: none"> • No missing outcome data; • Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); • Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardised difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size; • Missing data have been imputed using appropriate methods. <p>J.4.2 No (2) <i>Criteria for the judgement of 'NO' (i.e., high risk of bias).</i></p> <p><i>Any one of the following:</i></p> <ul style="list-style-type: none"> • Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardised difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size; • 'As-treated' analysis done with substantial departure of the intervention received from that assigned at
--	---

	<p>randomisation;</p> <ul style="list-style-type: none"> Potentially inappropriate application of simple imputation. <p>J.4.3 Unclear (2) <i>Criteria for the judgement of 'UNCLEAR' (uncertain risk of bias).</i></p> <p><i>Insufficient reporting of attrition/exclusions to permit judgement of 'Yes' or 'No' (e.g., number randomised not stated, no reasons for missing data provided).</i></p>
<p>J.5 Were the groups treated equally? <i>For example:</i></p> <p><i>(a) were the data collection measures for the intervention and control groups the same?</i></p> <p><i>(b) were the settings the same for both groups?</i></p> <p><i>(c) if relevant, was the activity delivered to both groups by the same person?</i></p> <p><i>(d) was there any relationship between the intervention and the outcome measures?</i></p>	<p>J.5.1 Not relevant (only one group) (2)</p> <p>J.5.2 Yes (1)</p> <p>J.5.3 No (2)</p> <p>J.5.4 Unclear (2) <i>Use this answer if relevant information is not reported.</i></p>
<p>J.6 Is/are the outcome measure(s) reliable?</p>	<p>J.6.1 Yes (1) <i>Select 'yes' if an instrument/questionnaire/test with reported reliability was used and/or where official datasets/standardised tests of attainment, e.g., GCSE records, were used.</i></p> <p><i>For example, standardised test for constructs such as motivation are coded 'yes'.</i></p> <p>J.6.2 No (2) <i>Select 'no' if an instrument/questionnaire without reported reliability was used, or a non-standardised test, or 'self-perception of impact' was measured.</i></p> <p>J.6.3 Unclear (2) <i>Use this answer if relevant information is not reported.</i></p>

Section K: Quality of the study (weight of evidence)

<p>K.1 Weight of evidence A: Taking account of all quality assessment issues, can the study findings be trusted in answering the study question(s)? <i>WOE A should be calculated from the scores in questions J.3, J.4, J.5 and J.6 (where 'yes'=1, 'no'=2, 'unclear'=2).</i></p> <p><i>The final score should be calculated as follows: (J.3 + J.4 + J.5 + J.6)</i></p> <p><i>High = 4 Medium/High = 5 Medium = 6 Low/Medium = 7 Low = 8</i></p> <p><i>Quality assessment is carried out at the level of an individual outcome, thus a study may have different quality scores for different outcomes.</i></p>	<p>K.1.1 High trustworthiness</p> <p>K.1.2 Medium/High trustworthiness</p> <p>K.1.3 Medium trustworthiness</p> <p>K.1.4 Low/Medium trustworthiness</p> <p>K.1.5 Low trustworthiness</p>
<p>K.2 Weight of evidence B: Appropriateness of research design and analysis for addressing the question, or sub-questions, of this specific systematic review. <i>Use the Maryland Scientific Methods Scale (MSMS) Score (see question H.1)</i></p> <p><i>High = MSMS score 5 (AND questions J.1 and J.2 must be answered 'YES'. If both not answered YES, then study to be judged as medium/high)</i> <i>Medium/High = MSMS score 4 (AND questions J.1 and J.2 must be answered 'YES. If both not answered YES, then study to be judged as medium)</i> <i>Medium = MSMS score 3</i> <i>Low/Medium = MSMS score 2</i> <i>Low = MSMS score 1</i></p>	<p>K.2.1 High</p> <p>K.2.2 Medium/High</p> <p>K.2.3 Medium</p> <p>K.2.4 Low/Medium</p> <p>K.2.5 Low</p>
<p>K.3 Weight of evidence C: Relevance of particular focus of the study (including conceptual focus, context, sample and measures) for addressing the question, or sub-questions, of this specific systematic review. In this review, all studies fixed at 'High'.</p>	<p>K.3.1 High</p> <p>K.3.2 Medium</p> <p>K.3.3 Low</p>
<p>K.4 Weight of evidence D: Overall weight of evidence <i>WOE D = whichever is the lowest score across WoE A and WoE B</i></p> <p><i>Quality assessment is carried out at the level</i></p>	<p>K.4.1 High</p> <p>K.4.2 Medium/High</p>

<i>of an individual outcome, thus a study may have different quality scores for different outcomes.</i>	K.4.3 Medium K.4.4 Low/Medium K.4.5 Low
---	---

Appendix 10: Studies excluded from the arts participation synthesis

The following studies were excluded from the arts participation synthesis as their overall quality of evidence (WoE D) was less than medium:

Bolduc J (2009) Effects of a music programme on kindergartners' phonological awareness skills. *International Journal of Music Education* 27: 37-47.

Kennedy R, Scott A (2005) A pilot study: the effects of music therapy interventions on middle school students' ESL skills. *Journal of Music Therapy* 42(4): 244-261.

O'Gara P (2008) To be or have not been: learning language tenses through drama. *Issues in Educational Research* 18(2): 156-166.

Schunk HA (1999) The effect of singing paired with signing on receptive vocabulary skills of elementary ESL students. *Journal of Music Therapy* 36(2): 110-124.

Appendix 11: Quality assurance (text mining)

In the final stages of the review process, additional steps were taken to check the accuracy of the text mining approach and our search strategy for identifying relevant studies. As this is a time-consuming activity, it could only be completed on a limited basis (with further work planned).

What did we do?

- Identify a recent review for each of the four sectors: arts, sport, libraries and museums (see box below);
- Scan the reference lists of these reviews for studies that answered our broad review question;
- Check to see if these items had (a) been identified by our original search (68,000 hits) and (b) were included in the sub-set of studies identified by text mining (12,439 hits).

Arts

In September 2009, an unpublished report by Professor Susan Hallam of London University's Institute of Education became available which presented an overview of empirical evidence on the impact of music on children and young people's intellectual, social and personal development (Hallam, 2009). Both UK and international research was included in the review.

Sport

A recent review conducted by a team of researchers from the Carnegie Research Institute examined participation in sport and physical recreation by black and minority ethnic (BME) communities (Long et al., 2009). The focus was (primarily) on UK literature published 1998-2009.

Libraries

Burns Owens Partnership (2009) undertook a short study (less than two months) to provide evidence of the type of data and research that is effective at capturing the impact of libraries on local communities and securing the support and engagement of key stakeholders. Both UK and international literature are included in the review.

Museums

Burns Owens Partnership (2005) conducted a review of past and current research into the social impact of museums (alongside that of libraries and archives). Literature from 1997 onward, both from the UK and overseas, was included in the review.