



A systematic review of sport and dance participation in healthy young people (15-24 years) to promote subjective wellbeing

Corresponding Author: Dr Louise Mansfield¹

Co-authors: Professor Tess Kay¹, Professor Catherine Meads², Dr Alistair John¹, Professor Norma Daykin³, Lily Grigsby Duffy¹, Jack Lane⁴, Professor Paul Dolan⁵, Stefano Testoni⁵, Professor Guy Julier⁴, Dr Annette Payne¹, Professor Alan Tomlinson⁴, Professor Christina Victor⁶

¹. Welfare, Health and Wellbeing, Institute for Environment, Health and Societies, Brunel University London.

². Anglia Ruskin University

³. University of Winchester

⁴. University of Brighton

⁵. The London School of Economics

⁶. Ageing Studies, Institute for Environment Health and Societies, Brunel University London

About the What Works Centre for Wellbeing

What Works Centre for Wellbeing is an independent organisation set up to produce robust, relevant and accessible evidence on wellbeing. We work with individuals, communities, businesses and government, to enable them to use this evidence to make decisions and take action to improve wellbeing.

The Centre is supported by the ESRC and partners to produce evidence on wellbeing in four areas: work and learning; culture and sport; community; and cross-cutting capabilities in definitions, evaluation, determinants and effects.

Culture and sport evidence team is comprised of researchers from:



Summary

We know that taking part in physical activity like sport and dance can bring wellbeing benefits, such as being more satisfied with life and happier, and feeling less anxious and depressed. Most of the evidence is however about adults. This review was carried out to investigate the relationships between subjective wellbeing (SWB) and taking part in sport and dance for healthy young people (15-24 years). Healthy people were defined as those without a condition diagnosed by a health professional. SWB describes wellbeing in terms of the good and bad feelings arising from what people do and how they think. The focus of this review was agreed through on-going collaborative engagement with UK-wide stakeholders representing policy, commissioning and managing, service delivery, and scholars from both academic and non-academic organisations.

We examined studies from the past 10 years and found that there is limited good quality evidence, and very little conducted in the UK. The review includes published findings from 977 participants across six countries - China, Korea, India, Turkey, Sweden and the USA. In some studies, participants were predominantly female. A wide variety of wellbeing measures were used. The most common form of sport/dance related activity was based on meditative practices (yoga and Baduanjin Qigong). Other physical activities reported included body conditioning, aerobic exercise, dance training, hip hop dance and sports including volleyball, ice skating, Nintendo Wii Active Games.

We included evidence from recent unpublished reports (grey literature) produced by or for sport and dance organisations since 2013. Participants in the evaluations were both male and female with a mean age between 13-24 years and were engaged in UK-based programmes of sport and dance. Findings illustrate that depending on activity type and delivery mode, taking part is associated with wellbeing improvements connected to social connectedness, pleasure, sense of purpose, confidence, interpersonal skills, happiness, relaxation, creative skills and expression, aspiration and ambition. Taking part was also associated with negative wellbeing connected to concerns about competency and capability.

Overall, the evidence available in this review suggests that yoga-type activities have the potential to improve subjective wellbeing and that group-based and peer supported sport and dance programmes may promote wellbeing enhancement in youth groups. The evidence in this review provides limited promising findings upon which sport and dance programmes for wellbeing improvement could be developed.

The lack of evidence identified in this review does not necessarily mean that wellbeing benefits are not accrued from taking part in sport and dance. There is scope to build evidence on wellbeing outcomes of sport and dance in healthy young people through well-designed, rigorous and appropriate research methods which are underpinned by relevant theory and use established methods of analysis.

Executive Summary

Introduction

The protocol for this review was registered on the PROSPERO International Prospective Register of Systematic Reviews (Registration number CRD42016048745).

The systematic review sought to address the question: What are the wellbeing outcomes of participation in sport and dance for healthy young people (15-24 years) and what are the processes by which wellbeing outcomes are achieved? Healthy young people were defined as those 15-24 years without a condition diagnosed by a health professional.

Review approach

The review included empirical research that assessed the relationship between sport and dance interventions with subjective wellbeing, published from 2006 – September 2016. Grey literature published from 2013-2016 was included.

Results

The electronic searches returned 6587 records for screening. Of these 8 were relevant studies of sport, dance and wellbeing in healthy young people.

3 evaluation reports were included in the grey literature from UK-based programmes of sport and dance for healthy young people.

Characteristics of included studies

The 8 quantitative peer reviewed and published studies investigated the effects of sport and dance interventions for a range of wellbeing outcomes. They include published data from 977 participants from 6 countries. Also included were 3 grey literature studies.

The published studies were 7 randomised control trials (RCTs) and 1 cohort study. There were some methodological challenges noted including small sample sizes, sample bias and limited analysis in some of the studies.

The most common form of intervention reported in three of the studies were based on meditative practices (yoga and Baduanjin Qigong). Other interventions reported included

body conditioning, aerobic exercise, dance training, hip-hop dance and sports including volleyball, ice skating, Nintendo Wii Active Games and an empowerment-based exercise intervention programme.

A wide variety of wellbeing measures were used in the published research and there was a great deal of heterogeneity across the studies. Meta-analysis was not appropriate in this systematic review.

Evaluation approaches used in the grey literature included surveys, focus groups, interviews and structured observations. Projects reported typically offered a wide range of sport, physical activity and dance activities to young participants including martial arts, dance, gym-based exercise, exercise classes, swimming, netball, cycling, and football, circus-based skills, and a range of dance forms. Interventions were led by instructors in group settings. Participants included both males and females with an average age of 13-24 years.

Summary of study findings

The strongest evidence (2 RCT studies, judged high quality) surrounds the effect of yoga or Baduanjin-Qigong on improving feelings of anxiety, depression, anger, attention and overall subjective wellbeing.

There is also moderate quality evidence to suggest that yoga can improve overall mood (1 study); that empowering young girls through exercise has a positive effect on self-efficacy (1 study); that aerobic and hip-hop dance lead to positive mood enhancement compared to ice-skating and body conditioning (1 study); and that dance training is effective in lowering self-reported depression (1 study).

There is low-moderate quality evidence (1 study) that a peer support exergaming programme can promote group cohesion and positive social reinforcement for taking part in physical activity in overweight young people.

There is low quality evidence (1 study) that ensuring positive feelings of competency, relationships with others and autonomy in competitive team sport (volleyball) players improves wellbeing and creates a more positive sport experience.

A small number of evaluation reports were included in the grey literature (n=3). Findings illustrate that depending on activity type and delivery mode, taking part is associated with wellbeing improvements connected to social connectedness, pleasure, sense of purpose, confidence, interpersonal skills, happiness, relaxation, creative skills and expression, aspiration and ambition. Taking part was also associated with negative wellbeing connected to concerns about competency and capability.

There was a high level of detail provided in the grey literature regarding evaluation methods. The strongest reports discussed evaluation approaches and methods of analysis in theoretical detail and acknowledged the limitations of evaluation design. Two studies reported both pre-project and post-project data. Regarding qualitative evaluation, all reports included collected rich data and provided in depth analysis. Data from the surveys tended to be relatively small scale.

Strengths and limitations of the review

The large number of hits following initial searches means that it is possible that some relevant evidence has not been included in this report. The focus on a specific target age group will have excluded evidence from studies that have aggregated data across younger and older age groups in their analysis. However, we undertook a comprehensive search strategy to identify all existing eligible studies published prior to the search dates. The pre-publication of our protocol on PROSPERO ensures methodological transparency and militates against potential post-hoc decision making which can introduce bias to the process. Dual screening of searches and data extraction and independent quality assessment using GRADE criteria ensured a rigorous process.

Taking published studies as the sole evidence increases the potential risk of publication lag, wherein possible important new evidence that has not yet been included in published reports is not identified and included. However, the grey literature review did include recent unpublished data from evaluations completed between 2013-2016. The quality of the grey literature and its focus on context-specific processes and outcomes gives them a high degree of credibility.

The use of the GRADE criteria introduces an element of subjective judgement. A consistent approach to judgements across the different interventions has been applied but it should be recognised that these judgements are open to interpretation.

Implications for research policy and practice

No UK studies were eligible for inclusion in this review. It is not possible to conclude that findings in this review are generalizable across countries. There is a need for national and local policy in sport and dance to recognise the potential significance of participation on subjective wellbeing and to support programmes that focus on building the evidence.

Recent national sport strategy in the UK (DCMS, 2015; Sport England, 2016) identifies wellbeing as an outcome for sport and physical activity and needs to be accompanied by attention to agreeing definitions and developing relevant (and potentially comparable)

measures of wellbeing outcomes, and evaluating what works to enhance wellbeing in sport and dance. National agencies may be influential in promoting this approach; conversely, a lack of national lead may discourage regional and local stakeholders from prioritising this.

The lack of evidence identified in this review does not necessarily mean that wellbeing benefits are not accrued from taking part in sport and dance. There is scope to build evidence on wellbeing outcomes of sport and dance in healthy young people through well-designed, rigorous and appropriate research methods which are underpinned by relevant theory and use established methods of analysis.

The development of a programme of wellbeing evaluation training would support key personnel in the sport and dance sectors in ensuring a comprehensive programme of delivery includes appropriate and rigorous monitoring and evaluation.

There is a need for studies of the wellbeing impacts of sport and dance for young people to be made public through academic and non-academic dissemination.

Introduction

Background

There is an established body of scientific evidence that increased levels of physical activity can bring wide-ranging physical and mental health benefits (Priest et al., 2008). Not taking enough physical activity is associated with an increased risk of a number of chronic diseases including coronary heart disease, diabetes, and some forms of cancer (Foster et al., 2005). Physical activity is also associated with the effective risk reduction of developing and living with mild to moderate mental health conditions such as depression and anxiety (Start Active Stay Active, 2011; WHO, 2010). In the UK the Moving More, Living More cross-government group recognises the role that sport can play in helping people to become more active for health and wellbeing benefits (Mansfield et al., 2015). Dance, in various forms, is also being promoted for health and wellbeing enhancement (OneDanceUK, 2017). Both sport and dance organisations identify young people as a key target group for engagement in physical activity to enhance, health, wellbeing and personal development.

There is emerging evidence of the association between sport-related activities that could include dance in achieving mental health and wellbeing outcomes (Fujiwara et al., 2014). Research also highlights the need for further understanding of social diversity and context in evaluating and understanding subjective wellbeing and this is significant to community sport and dance (Dolan and Galizzi, 2015). Overall, the extant literature is theoretically and methodologically diverse and less attention has been given to children and adolescents. Existing evidence reviews on sport have tended to focus on physical rather than mental health outcomes (see for example Cavill et al., 2012; Oja et al., 2011; Oja et al., 2015) or they have examined the effect of exercise in populations with specific mental health conditions such as depression (Mammen and Faulkner, 2013) and anxiety (Ströhle, 2009). Dance-related reviews of evidence have examined the effectiveness of dance therapy on psychological and physical outcomes in cancer patients (Bradt et al., 2015), for schizophrenia (Ren and Xia, 2013) and on depression (Meekums et al., 2015). A review of reviews on physical activity and mental health in children and adolescents identifies an association between physical activity and positive outcomes connected to depression, anxiety, self-esteem and cognitive function (Biddle and Asare, 2011). To our knowledge no systematic review to date has focused on sport and dance participation in healthy young people (15-24 years) to promote subjective wellbeing.

There is evidence that community engagement interventions can have a positive impact on health and wellbeing (O'Mara et al., 2013; South, 2015). Sport, in the UK has traditionally been delivered through formal clubs, and dance is also often accessed via formalised school

or class provision. Such delivery models usually require a financial commitment to membership, involve the development of skills and performance characteristics, and are commonly based on one sport form or dance genre. One of the current concerns in the sport sector particularly but also arising in relation to dance, is the role of traditional club-based systems in increasing participation versus the development of more informal models of delivery (e.g. parkour, skateboarding, park run) which have developed outside the structures of traditional sport (see for example Wheaton et al., 2015; Wheaton and Gilchrist, 2017 forthcoming).

The aim of this systematic review is to firstly evaluate the subjective wellbeing outcomes in healthy young people who participate in sport and/or dance activities in club and non-club contexts compared to inactive controls. Secondly, we seek to establish if the informal aspect of sport or dance participation is more likely to lead to wellbeing enhancement than participation in club-based sport and dance.

The protocol for this review was registered on the PROSPERO International Prospective Register of Systematic Reviews (Registration number CRD42016048745).

Research Question

What are the wellbeing outcomes of participation in sport and dance for healthy young people (15-24 years) and what are the processes by which wellbeing outcomes are achieved?

Methods

Types of studies

We included published studies that assessed the relationship between sport and dance interventions and subjective wellbeing in young healthy people i.e. with no diagnosed chronic disease. One study recruited overweight and obese participants diagnosed by a health professional but the sample is predominantly in the overweight category which (unlike obesity) is not a diagnosed health condition and we have included it. Some studies included measures of depression, however, the participants were not diagnosed with depression and were considered to be healthy and we have included them. We included empirical research: quantitative, qualitative or mixed methods, outcomes or process evaluations, published from 1996 – September 2016. We identified systematic reviews published between 2010 and 2016 for the purposes of hand searching the reference lists. Grey literature published 2013-2016 was included.

Types of participants

We included healthy young people (15-24 years old). Healthy people were defined as those without a condition diagnosed by a health professional. This included any group or individual taking part in performing or watching sport or dance type activities but not as paid professionals or training to be an elite paid performer, and not in clinical sport-based or dance therapy. We intended to include studies from countries economically similar to the UK (i.e. other high income countries with similar economic systems). Countries in which the studies are based are listed in table 1. All except India and Korea are categorised in the same group as the UK in the OECD Development Assistance Committee categories. The sample participants in the studies conducted in India and Korea are University students and likely to be relative high in socioeconomic status and we have included them here.

Types of outcome measure

In order to be included, studies needed to have measured subjective wellbeing using any recognised method or measure. A summary of the wellbeing measures used in the studies included in this review can be found in Appendix 1. For the health economic component key outcomes were the outputs from cost, cost-utility, cost-effectiveness, cost-benefit and cost-consequence analyses.

Types of interventions

We focused on participatory sport and dance interventions including watching and performing. We included sport-related and dance therapy offered to enhance wellbeing in healthy young people. We excluded evidence relating to paid professional sports and dance people and clinical sport-based or dance therapy.

Comparison

No sport or dance, or usual routine i.e. inactive comparator or historical/time-based comparator.

Search methods for identification of reviews

Electronic searches

Electronic databases were searched using a combination of controlled vocabulary (MeSH) and free text terms. Search terms were incorporated to target empirical evidence on music, singing and wellbeing (Figure 1). We incorporated specific filters to identify health economic evaluations. The OVID MEDLINE search strategy can be found below. All database searches were based on this strategy but were appropriately revised to suit each database. The following databases were searched from 1996-2016:

- PsychInfo
- OVID MEDLINE
- Eric
- Arts and Humanities Citation Index (Web of Science)
- Social Science Citation Index (Web of Science)
- Science Citation Index
- Scopus
- PILOTS
- CINAHL
- SportDiscuss
- International Index to Performing Arts (IIPA)

For the review of health economic evaluations we will separately search the following databases

- OVID MEDLINE
- Scopus
- CINAHL
- NHS EED (NHS Economic Evaluation Database)
- HTA Technology Assessment) database

Search Strategy (OVID MEDLINE)

1. MeSH descriptor: [well being]
2. well-being
3. wellbeing
4. "young people".mp or youth.mp or adolescent*.mp
5. sport/ or sport.mp.
6. "physical activity".mp or "physical activity"/
7. Exercise*.mp.
8. "physical exertion".mp.
9. dance*.mp.
10. game*.mp.
11. team.mp.
12. bike.mp.
13. cycl*.mp.
14. cheerlead*.mp.

15. equestrian.mp.
16. swim*.mp.
17. gym* .mp.
18. sail*.mp.
19. canoe*.mp.
20. kayak*
21. bloodsport*.mp.
22. boxing.mp
23. "martial arts".mp.
24. fitness.mp.
25. ballet.mp.
26. choreograph*
27. "work-out".mp.
28. (1 or 2 or 3) and (4) and (or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18, or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27)
29. tournament.mp
30. match.mp
31. competition.mp
32. festival.mp
33. battle.mp
34. league.mp
35. team*.mp
36. theatre*.mp
37. event*.mp
38. meet*.mp
39. field*.mp
40. fan.mp
41. play*.mp
42. athlet*.mp
43. attend*.mp
44. spectat*.mp
45. particpat*.mp
46. perform*.mp
47. 28 and (29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46)
48. Quality of life.mp. or "Quality of Life"/ Life
49. Anxiety/ or anxiety.mp.
50. self-esteem.mp.
51. loneliness/ or lonel. mp.
52. life adj satisfaction.mp.
53. happiness.mp.
54. worthwhileness.mp.
55. 47 and (48 or 49 or 50 or 51 or 52 or 53 or)
56. limit 54 to humans and all young people or adolescents

Searching other sources

The reference lists of all relevant reviews from the last 5 years were hand-searched to attempt to identify additional relevant empirical evidence. A search of UK grey literature was conducted via an online call for evidence, employment of expert input, review of key sector websites and a Google search. Grey literature was included if it was a final evaluation or report on empirical data, had the evaluation of sport-related or dance interventions as the central objective, was published 2013-2016, and included details of authors (individuals, groups or organisations).

Identification of studies for inclusion

Search results were independently checked by two review authors. Initially the titles and abstracts of identified studies were reviewed. If it was clear from the title and abstract that the study did not meet the inclusion criteria it was excluded. Where it was not clear from the title and abstract whether a study was relevant the full article was checked to confirm its eligibility. The selection criteria were independently applied to the full papers of identified reviews by two review authors. Where two independent reviewers did not agree in their primary judgements they discussed the conflict and attempted to reach a consensus. If they could not agree then a third member of the review team considered the title and a majority decision was made. Studies in any language were included. A table of excluded studies can be found in Appendix 2.

Data collection and analysis

Data extraction and management

Data were extracted independently by two review authors using a standardised form (Appendix 3). Discrepancies were resolved by consensus. Where agreement could not be reached a third review author considered the paper and a majority decision was reached.

For quantitative evidence of intervention effectiveness, the data extraction form included the following details:

- evaluation design and objectives (the interventions studied and control conditions used, including detail where available on the intervention content, dose and adherence, ethics)
- sample (size, representativeness, reporting on drop-out, attrition and details of participants including demographics and protected characteristics where reported)

- the outcome measures (the scales used and the collection time-points, independence, validity, reliability, appropriateness to wellbeing impact questions)
- analysis (assessment of the methodological quality/risk of bias)
- results and conclusions
- the presence of possible conflicts of interest for authors

For qualitative evidence of intervention effectiveness the data extraction form included the following details:

- research design and objectives (interpretive, examining subjective experiences of participants, ethics)
- data collection (type/form, appropriateness, recording, theoretical justification)
- participants (numbers and details including demographic, recruitment strategy, theoretical justification)
- analysis (rigor, assessment of methodological quality, identification of bias/involvement of researcher, attribution of data to respondents, theoretical justification, relevance to wellbeing impact question)

Health economic studies will be extracted when there is agreement by WWW evidence review programmes regarding methods. We do not report on health economic studies here. Our approach would be to extract the following additional information:

- Included study designs, analytic methods, perspective, time horizon, discount rate
- type of sensitivity analysis undertaken
- type and sources of data use for resource use and costs, reporting figures for costs;
- methods of preference elicitation (e.g. contingent valuation, revealed preferences, trade-off methods), reporting estimates of preference values
- main results including specified types of incremental cost-effectiveness ratios - ICERs(e.g. health service or societal perspective)
- main health economic conclusions of the review

Our protocol allowed us to contact the authors of articles in the event that the required information could not be extracted from the studies if this was essential for interpretation of their results. We did not need to take this action.

Assessment of methodological quality of included studies

We used the quality checklists for quantitative and qualitative studies detailed in the What Works Centre for Wellbeing [methods guide](#), and for economic evaluations (The Drummond Checklist, 1996) to assess the methodological quality of the included studies.

Included studies assessed the methodological quality/risk of bias in a variety of ways. We refer to the judgements made by the authors of studies regarding the quality of evidence/risk of bias and report it within the context of our assessment of the quality of a study itself. We employed GRADE schema for judging certainty / quality of evidence as high quality, moderate quality, low quality and very low quality.

Results

Results of the searches

After removal of duplicates the electronic searches returned 5597 records for screening. Of these, 149 were retained after abstract and title screening and the full texts were assessed. 141 records were excluded at this stage and relevant records were identified through hand-searching of the reference lists of included reviews, resulting in 8 records in total. The search screening process is illustrated in Figure 1.

The search identified 8 published studies on sport, dance and wellbeing in young people (aged 15-24 years). The list of included studies can be found in the references section and Table 1.

Characteristics of Included Studies

The included studies investigated the effects of different types of sport and dance interventions for a wide range of wellbeing outcomes. Sport and dance are defined broadly in this evidence programme and include all forms of physical activity; formally or informally organised, and involving individual or group participation, for expressing or improving physical fitness or forming social relationships or for competition. The review does not include walking interventions as there are several existing reviews of walking (see for example Ogilvie et al., 2007; Ogilvie et al., 2004)

Three surveys with 2 or more data collection time points met the inclusion criteria but it was agreed to exclude these as no specific sport or dance interventions were identified. The surveys identified correlations between health enhancing physical activity and wellbeing, however, the Culture and Sport Evidence Programme includes a secondary analysis of largescale data sets which will give the most up-to-date evidence on the relationships between sport, dance and healthy young people.

A summary of the characteristics of the included papers is presented in Table 1.

Overview of Quality of Included Studies

The review includes 7 randomised control trials and 1 cohort study (pre/post-test design). There were some methodological challenges noted including small sample sizes, sample bias and limited analysis in some of the RCT studies, and an absence of qualitative studies in the published literature.

The use of the GRADE schema for judging quality of evidence resulted in 2 high quality RCT studies, 4 moderate quality RCT studies, 1 low quality RCT and 1 low quality cohort study. There are a small number of published studies examining the subjective wellbeing outcomes of sport and dance for healthy young people (aged 15-24 years). Despite the 2 high quality RCTs, overall the quality is moderate-low in respect of there being very little evidence in total and that the high quality RCTs are small (n=222 and n=50), include sample bias and limited analysis in some cases.

Grey Literature Searches

The grey literature search was undertaken concurrently with the Culture and Sport systematic review on sport, dance, healthy young people and wellbeing. A call for UK grey literature evidence of wellbeing impacts of sport and dance interventions with young healthy people was placed on the What Works Wellbeing website between October and November 2016. Our focus was on collecting, reviewing and synthesising most recent work in the sector in order to assess current approaches. The call requested evaluation reports completed between 2013 and 2016. Additionally, we conducted an extended systematic search of grey literature by employing expert input that assisted in identifying sources of grey literature that might not be readily available in searching peer-reviewed literature (Benzies et al., 2006). Specifically, we (i) contacted known experts in the field for recommendations of sport or dance sector repositories (ii) reviewed websites of key sport and dance organisations (iii) searched the EThOS website for unpublished PhD dissertations and (iv) conducted a Google search with key words - sport, dance, wellbeing, young people and reviewing titles of first 100 hits.

A total of 54 submissions were screened by the research team, of which 3 met the inclusion criteria. Submissions reviewed for eligibility included 12 received through the call for evidence, 33 obtained via the extended search for grey literature and 9 PhDs. Reasons for exclusion were 'not sport or dance intervention', 'not healthy young people', 'not published between 2013-2016', and / or 'not wellbeing related'. Table 2 presents a summary of the grey literature (3 evaluation reports) included in this review.

In order to capture project details we used an adapted version of the Public Health England Arts and Health Evaluation Framework (Daykin with Joss, 2016) to record information such as project activity, aims, location, setting, timescale, population, costs and reported outcomes. We also recorded evaluation details where reported including rationale, method, costs, data collection and analysis techniques, and findings.

Projects reported in the grey literature typically offered a wide range of sport, physical activity and dance activities to young participants. Evaluation approaches used in the reports included surveys, focus groups, interviews and structured observations. There was a high level of detail provided regarding evaluation methods which were systematic. The strongest reports discussed evaluation approaches and methods of analysis in theoretical detail and acknowledged the limitations of evaluation design. Two studies reported both pre-project and post-project data. Regarding qualitative evaluation, all reports included collected rich data and provided in depth analysis. Data from the surveys tended to be relatively small scale. Whilst this might limit the strength of the evidence and make it difficult to generalise from the findings, the quality of the evaluations and their focus on context specific processes and outcomes gives them a high degree of credibility.

Figure 1: PRISMA flow diagram of the search screening process

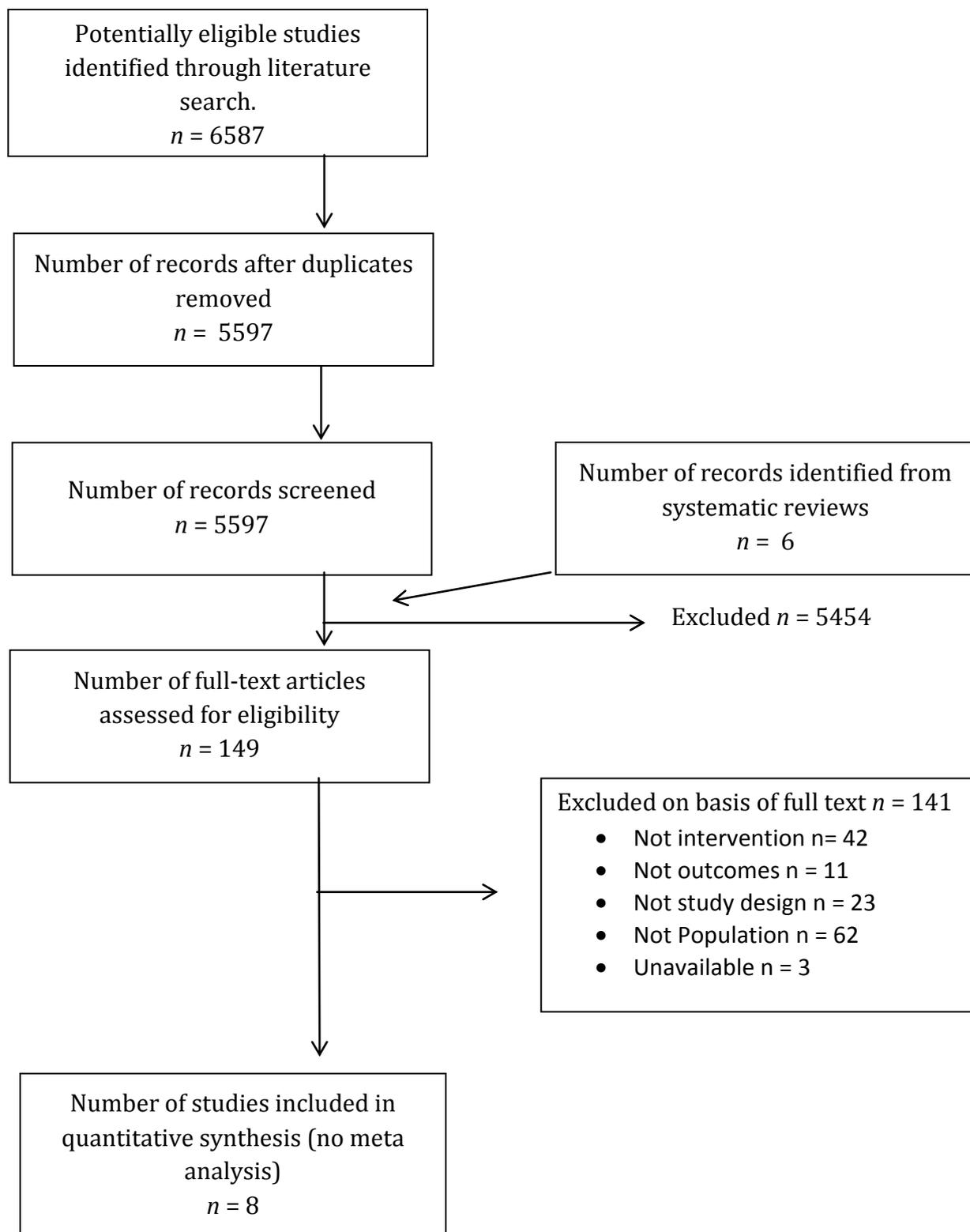


Table 1. Characteristics of included studies (studies on healthy young population aged 15-24 years)

Authors (Country)	Date	Numbers of Participants (recruited and completed)	Participant Description (include protected characteristics)	Intervention/ comparison (include a short description of the intervention)	Outcomes and measures used	Study Design	Conclusions	Limitations (risk of bias)
Akandere and Demir (Turkey)	2011	120	healthy male (50%) and female conservatory student volunteers -aged 20 - 24	12 week dance training intervention	Depression (Beck Depression Scale) before and after 12week dance intervention	RCT	Dance had a positive effect over the depression levels at the end of 12 weeks of dance training (p<0.05)	<ul style="list-style-type: none"> • Only one measure used • Small population • sample already had dance knowledge • participant details not clearly reported • baseline levels of depression differ in groups
Amorose et al (USA)	2009	93	female adolescent volleyball players. Age: 13-18 (M=15.78 yrs). Mostly Caucasian (90.6%).	Followed a cohort of female adolescent volleyball players through a season of competitive volleyball games (approx. 4 months)	1. Need Satisfaction -5 item Subscale of the Intrinsic Motivation Inventory (assess sport competence) -6 item Scale (Hollembek & Amorose 2005) (assess need for autonomy) -10 item Richer &	Cohort	the bivariate correlations at both preseason and postseason time points were consistent with Basic Needs Theory in that need satisfaction related positively to self-esteem and negatively to burnout. "Changes in athletes'	<ul style="list-style-type: none"> • sample bias: one club in Western U.S., one sport. All females. Mostly Caucasian • Selection bias:

					<p>Vallerand's Feelings of Relatedness Scale (assess need for relatedness)</p> <p>2. Well being</p> <p>-10 item Rosenberg's Self-Esteem Scale (measuring self-esteem)</p> <p>-15 item Athlete Burnout Questionnaire (measuring burnout)</p> <p>Administered to pts pre (1-2 wks before competitive season starts) and post season (1-2 wks before the last official game / ~4m after start of season)</p>		<p>need satisfaction over the course of their season related to changes in their overall wellbeing"</p> <p>"The extent to which athletes' psychological needs are satisfied during a season are linked to increases and decreases in their positive and negative wellbeing."</p>	<p>only those that agreed to volunteer. drop out not reported</p> <ul style="list-style-type: none"> Study design: no control group. Only 2 time points looked at did not assess social contextual factors e.g. coaching behaviour
Kanojia et al (India)	2013	50 (25 in intervention, 25 in control)	healthy female volunteers	<p>Yoga practiced for 35-40 minutes/day x 6 days/week for the duration of 3 menstrual cycles. Training given by qualified instructor vs no intervention</p>	<p>-Anger (16 item questionnaire)</p> <p>-Trait anxiety (40 item questionnaire)</p> <p>-Depression (10 item questionnaire)</p> <p>-Subjective well-being (50 item questionnaire)</p> <p>Questionnaires were developed by DIPAS (Defense Institute of Physiology and Allied Sciences)</p>	RCT	<p>- significantly higher anxiety, depression and anger scores and decrease in sense of well-being score during premenstrual phase compared with postmenstrual phase in all the three cycles...</p> <p>With the practice of yoga, significant decrease in anxiety, depression and anger scores and improvement in sense of well-being was recorded from initial to second and third menstrual cycle both in premenstrual and postmenstrual phases.</p>	<ul style="list-style-type: none"> Drop out not reported recruitment methods not reported Not possible to double blind <i>Consistent findings</i>

							No significant change was observed in control group.	
Kim & Kim (Korea)	2007	277	Korean high school (n = 45) and undergraduate students (n = 232) volunteers. Age range: 17 - 22 (mean 20.6 (SD = 1.2))	1 of 4 40 minute exercise sessions: aerobic exercise, body conditioning, hip-hop dancing, and ice skating	Mood (Subjective Exercise Experiences Scale: measuring 3 dimensions; positive well-being, psy distress, and fatigue) before and after exercise session.	RCT	aerobics and hip-hop dancing groups rated positive well-being higher than the body conditioning and ice skating groups and rated lower psychological distress and fatigue.	<ul style="list-style-type: none"> Data based on one session only.
Li et al (China)	2015	222	college students recruited from the first or second grade in FJTCM intervention: n=101 (85.1% female Mage: 20.63) Control: n=101 (80% female Mean age: 20.92)	Baduanjin exercise 1hr/day 5x week x12 weeks	<p>Self-esteem – SES Scale (Higher scores = higher self-esteem)</p> <p>Mood / mindfulness POMS scale (Higher scores = more negative current mood states)</p> <p>QoL – WHOQOL-BREF</p> <p>Attention – Schulte Grid (Less time represents higher level of attention)</p> <p>Stress - CPSS Scale</p>	RCT	Compared with controls, significant improvements in Baduanjin exercise group at the end of 12-week intervention period were found on attention. No significant changes in other mental outcomes, including self-symptom intensity, stress, self-efficacy, quality of life, and quality of sleep, were found between groups. No adverse event was reported during the study period	<ul style="list-style-type: none"> not blinded participants recruited from one medical university greater proportion of female participants small effect size <i>excellent protocol adherence</i> <i>no significant loss to follow-up</i>
Lindgren et al (Sweden)	2010	110 recruited (54 intervention, 56 control) 62 completed (27	Pts from secondary schools in low SES areas. 100% female Mean age: 15.3 (intervention), 15.5 (control)	Invited to participate in different sports and exercise activities aimed at empowering pts. 2x wk for 6 months vs waiting list (control)	<p>Self-efficacy - Swedish version of a 10-item General Self-efficacy Scale</p> <p>Social Barriers to Exercise Self-efficacy Questionnaire</p>	RCT	Sports and exercise had an impact on adolescent girls' general perceived self-efficacy. This particular result can be considered as an empowering effect	<ul style="list-style-type: none"> small sample size high dropout rate

		intervention, 35 control)	All physically inactive				because participants reported an increased, stable sense of personal ability to deal effectively with a variety of stressful situations in general. Other effects of were the significant changes within the intervention group in terms of physical fitness level.	
Noggle, J. et al (USA)	2012	51 recruited, 36 in intervention, 15 in control.	<p>Grades 11 and 12 PE class at a public high school</p> <p><u>Intervention</u> N = 36, 61% female, Age (SD) 17.1 (0.6).</p> <p><u>Comparator</u> N = 15, 47% female, Age (SD) 17.3 (0.8).</p>	A Kripalu-based yoga program of physical postures, breathing exercises, relaxation, and meditation was taught 2 to 3 times a week for 10 weeks (28 yoga sessions total). Self-report questionnaires were administered to students 1 week before and after.	<p>Primary outcome measures of psychosocial well-being; Profile of Mood States (SF); Positive and Negative Affect Schedule for Children. Additional measures of psychosocial well-being; Perceived Stress Scale and Inventory of Positive Psychological Attitudes. Secondary measures of self-regulatory skills included Resilience Scale, State Trait Anger Expression Inventory-2TM, and Child Acceptance Mindfulness Measure.</p>	RCT	<p>Implementation of the yoga class was feasible and students generally found it beneficial. Although not causal due to small, uneven sample size, this preliminary study suggests preventive benefits in psychosocial well-being from Kripalu yoga during high school PE.</p>	<ul style="list-style-type: none"> • Small sample size. • would have been ideal to randomise individually but being in a school setting required allocation at the classroom level • Moderate attendance at the yoga classes
Staiano, A. et al (USA)	2013	N = 54 <i>Competitive</i> exergame (EG) (baseline n=19, time 2	54 African American adolescent's aged 15-19 years (55.6% female). Mean participant BMI	All EG participants were encouraged to play the Nintendo Wii Active game for 30-60 minutes per school day in a lunch-time or	<p>Self-reports of:</p> <ul style="list-style-type: none"> - Self-efficacy (Exercise Confidence Survey) - Self-esteem (Rosenberg Self-Esteem scale) - Peer support (Friendship 	RCT	(Growth curve analysis) Co-op. EG players lost significantly more weight ($M = 1.65$ kg; $SD = 4.52$) than controls (no weight loss). Comp. EG players no	<ul style="list-style-type: none"> • Sample bias: small sample from one school and some attrition

	<p>n=19, time 3 n=17). <i>Cooperative EG</i> (baseline n=19, t2 n=19, t3 n=10). <i>Control</i> (baseline n=16, t2 n=16, t3 n= 12)</p>	<p>percentile at baseline was 94.7 (SD=6.0). 95% is regarded as obese. Above 85% is overweight.</p>	<p>after-school program. Cooperative EG participants worked with a peer to expend calories and earn points together, whereas competitive EG participants competed against a peer.</p>	<p>Quality Questionnaire)</p>	<p>sig. difference from other conditions. Co-op EG players sig. increased self-efficacy over controls. Both EG conditions significantly increased in peer support (more than controls). EG's played cooperatively can be effective technology for weight loss among youth.</p>
--	---	---	---	-------------------------------	--

Table 2. Summary of included grey literature

Author (year)	Submission Title	Organisation	URL
Susan Potter, Filipa Pereira Stubbs (2015)	DanceQuest 2012-15 Evaluation Report	DanceQuest: Children & the Arts	URL Not available
BOP Consulting (2016)	The Social Impact of Jacksons Lane	Jacksons Lane	https://jacksonslane.rooftop.io/sites/18/2016/10/05112305/JL-Social-Impact-Study.pdf
Louise Mansfield, Tess Kay, Nana Anokye, Julia Fox-Rushby, (2016)	The Health and Sport Engagement (HASE) intervention and evaluation project (2013-2016): the design, delivery and evaluation of a complex community sport intervention for improving physical activity, health and wellbeing. Volume 1: Evaluation of the Design and Delivery of the HASE project	Brunel University	URL not available. To be published in April 2017. Under review with Sport England

Findings of included papers

Study participants

The review includes published data from 977 participants from six countries - China, Korea, India, Turkey, Sweden and the USA. Of the total participants 884 were involved in randomised controlled designs and 93 in a cohort study. Studies included male and female young people with a mean age between 15-24 years old although data show that young women were predominantly recruited to some studies. Where demographic characteristics of participants were reported, this revealed a mix of ethnic backgrounds including white Caucasian, Chinese, Indian, Korean, and African American. In the grey literature participants were both male and female with a mean age between 13-24 years and were engaged in UK-based programmes of sport and dance. Grey literature reporting a younger mean age was included if they revealed particularly informative findings for the evidence review and where the evidence from published studies was limited.

Types of sport and dance interventions

In the published literature

The most common form of intervention reported were based on meditative practices including yoga (Kanojia et al., 2013; Noggle et al., 2012) and Baduanjin Qigong (Li et al., 2015). Other interventions reported included body conditioning (Kim and Kim, 2007) aerobic exercise (Kim and Kim, 2007), dance forms delivered through dance training (Akandere and Demere 2011), hip-hop dance (Kim and Kim, 2007), an empowerment-based exercise intervention programme (Lindgren et al., 2010) and specifically identified sports including ice skating (Kim and Kim 2007) and Nintendo Wii Active Games (Staiano et al., 2013). The cohort study followed volleyball players through a season of competitive games (Amorose et al., 2009).

Interventions in 6 of the RCT studies and in the cohort study were led by sport or dance instructors in formal group sessions. 1 RCT used the Nintendo Wii Active Games Programme incorporating a cooperative peer-to-peer method of participation.

In the grey literature

A variety of UK-based sport, physical activity and dance activities were evaluated and reported on in the grey literature. Interventions included martial arts, dance, gym-based

exercise, exercise classes, swimming, netball, cycling and football (Mansfield et al., 2016), circus-based skills (e.g. juggling, balancing, diabolo) (BOP Consulting, 2016), and a range of dance forms (Potter et al., 2015). Interventions were led by instructors in group settings.

Wellbeing measures

A wide variety of wellbeing measures were used in the published literature, perhaps reflecting the fact that *subjective wellbeing* is a relatively recent topic of study in the sport and dance sectors and that associated concepts such as self-esteem, confidence and anxiety have historically provided an emphasis for measurement. In the UK, it is only since April 2011 that personal wellbeing has been measured by the ONS. The Annual Population Survey (APS) includes four questions which are used to monitor personal wellbeing in the UK; (1) Overall, how satisfied are you with your life nowadays? (2) Overall, to what extent do you feel the things you do in your life are worthwhile? (3) Overall, how happy did you feel yesterday? (4) Overall, how anxious did you feel yesterday?

The measures used in the included published studies reflect some of these domains to some extent. An overview of outcome measures used in the included studies can be found in Appendix 1. One study utilised a satisfaction with life scale (SLWS Diener et al., 1985) and one other measured quality of life (WHOQOL-BREF scale). Other validated measures encompassed several dimensions of psychological wellbeing, such as self-esteem, emotion, enjoyment, purpose in life, burnout, confidence, stress, anger, anxiety, depression, hedonic wellbeing (through increased pleasure / decreased pain) and eudemonic wellbeing (through meaning and self-realization), relationships with others and mood. A summary of the numerical results from the included studies can be found in Table 3.

In the grey literature, surveys, focus groups, interviews and structured observations evaluated wellbeing using descriptive statistics and/or thematic analysis. Evaluations discussed subjective wellbeing in terms of social connectedness, feelings of pleasure, the development of a sense of purpose, exhilaration, confidence, self-esteem, (happiness, relaxation, playfulness, fun), gaining creative skills and expression, combatting isolation, improved sense of being a part of something positive, and increased aspiration and ambition.

Table 3 Summary of numerical results (published studies)

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
Akandere, M. and Demir, B. (2011)	12 week dance training intervention	Depression (Beck Depression Scale)	N =60 15.72 (7.004)	N= 60 16.53 (5.922)	N= 60 13.90 (5.568) ¹	N= 60 17.48 (7.740) ²	N/A	N/A	<ul style="list-style-type: none"> • Only one measure used • Small population • sample already had dance knowledge • participant details not clearly reported • baseline levels of depression differ in groups
Amorose, A.J., Anderson-Butcher, D. and Cooper, J.	A season of competitive volleyball games (approx. 4 months)	-Need satisfaction; sport competence, need for autonomy,	N=93 NR		N=93 NR		N/A	N/A	<ul style="list-style-type: none"> • sample bias: one club in Western U.S., one sport. All

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
(2009)		need for relatedness. -Self-esteem (10 item Rosenberg's Self-Esteem Scale) -Burnout (15 item Athlete Burnout Questionnaire)							females. Mostly Caucasian <ul style="list-style-type: none"> • Selection bias: only those that agreed to volunteer. drop out not reported • Study design: no control group. Only 2 time points looked at • did not assess social contextual factors e.g. coaching behaviour

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
Kanojia, S., Sharma, V.K., Gandhi, A., Kapoor, R., Kukreja, A. and Subramanian, S.K. (2013)	Yoga practiced for 35-40 minutes/day x 6 days/week for the duration of 3 menstrual cycles.	Anger (16 item questionnaire)	N=25 Postmenstrual phase: Initial cycle 8.84 (4.01), Premenstrual phase: Initial cycle 15.0(5.92)###	N=25 Postmenstrual phase: Initial cycle 9.12(4.41), Premenstrual phase: Initial cycle 14.32(5.24)###	N= NR Postmenstrual 2nd cycle 7.76 (3.53)***, Premenstrual 2nd cycle 9.52 (4.70)***##	N= NR Postmenstrual 2nd cycle 9.04(4.33), Premenstrual 2nd cycle 14.28(4.89)###	N= NR Postmenstrual 3rd cycle 7.92 (4.29) Premenstrual 3rd cycle 8.52 (4.15)***+	N= NR Postmenstrual 3rd cycle 8.96(4.65) Premenstrual 3rd cycle 13.12(4.83)###	<ul style="list-style-type: none"> • Drop out not reported • recruitment methods not reported • Not possible to double blind • <i>Consistent findings</i>
		Trait anxiety (40 item questionnaire)	N=25 Postmenstrual phase: Initial cycle 40.64 (6.22), Premenstrual phase: Initial cycle 46.96 (5.87)###	N=25 Postmenstrual phase: Initial cycle 41.6(5.49), Premenstrual phase: Initial cycle 46.76(5.33)###	N=NR Postmenstrual 2nd cycle 39.40 (6.69), Premenstrual 2nd cycle 41.48 (5.77)***#	N= NR Postmenstrual 2nd cycle 40.24(6.97), Premenstrual 2nd cycle 45.80(6.41)###	N= NR Postmenstrual 3rd cycle 37.24 (9.14)*+ Premenstrual 3rd cycle 40.80 (5.75)***#	N= NR Postmenstrual 3rd cycle 38.64 (12.76) Premenstrual 3rd cycle 43.88(7.06)	
		Depression (10 item questionnaire)	N=25 Postmenstrual phase: Initial cycle 6.84 (3.10) Premenstrual phase: Initial cycle 10.72 (4.19)###	N=25 Postmenstrual phase: Initial cycle 6.36(4.13), Premenstrual phase: Initial cycle 9.72(3.89)##	N= NR Postmenstrual 2nd cycle 3.96 (2.59)*** Premenstrual 2nd cycle 5.92(3.76)***##	N= NR Postmenstrual 2nd cycle 6.24(4.98), Premenstrual 2nd cycle 9.56(3.22)##,	N= NR Postmenstrual 3rd cycle 3.12 (2.71)***++ Premenstrual 3rd cycle 4.76(2.82)***+ +###	N= NR Postmenstrual 3rd cycle 6.07(2.81) Premenstrual 3rd cycle 9.36(2.96)##	
		Subjective well-being (50 item questionnaire)	N=25 Postmenstrual phase: Initial cycle 41.72	N=25 Postmenstrual phase: Initial cycle	N= NR Postmenstrual 2nd cycle 39.64(16.07)**,	N= NR Postmenstrual 2nd cycle 44.68(16.5),	N= NR Postmenstrual 3rd cycle 37.20(15.17)**	N= NR Postmenstrual 3rd cycle	

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
		re)	(16.05), Premenstrual phase: Initial cycle 53.92 (20.35)###	45.6(14.05), Premenstrual phase: Initial cycle 51.04(14.89),	Premenstrual 2nd cycle 44.48 (17.87)***##,	Premenstrual 2nd cycle 50.40(18.67),	*++ Premenstrual 3rd cycle 40.24 (16.22)***+++	43.96(14.01) Premenstrual 3rd cycle 49.76(17.02)#	
Kim, S. and Kim, J. (2007)	1 of 4 40min exercise session: aerobic exercise, body conditioning , hip-hop dancing, and ice skating	Positive wellbeing (Subjective Exercise Experiences Scale)	Aerobic dance (n=84): 16.8 (4.0) Body conditioning (n=64): 15.3 (2.9) Hip-hop dance (n=45): 16.3 (4.2) Ice skating (n=84): 19 (3.9)		Aerobic dance (n=84): 19.9 (3.9) ¹ Body conditioning (n=64): 18 (2.8) ² Hip-hop dance (n=45): 19.7 (3.4) ¹ Ice skating (n=84): 20.4 (3.4) ²		n/a	n/a	• Data based on one session only
		Psychological distress (Subjective Exercise Experiences Scale)	Ice skating: 8.3 (3.9) Hip-hop dance: 9.8 (4.6) Body conditioning: 10.7 (4.1) Aerobic dance: 9.4 (4.2)		Ice skating: 8.1 (3.9) ² Hip-hop dance: 7.3 (4.2) ¹ Body conditioning: 9.6 (3.2) ² Aerobic dance: 6.7 (2.9) ¹		n/a	n/a	
		Fatigue (Subjective Exercise Experiences Scale)	Ice skating: 10.9 (5.4) Hip-hop dance: 16.2 (4.4) Body conditioning: 15.9 (4.4) Aerobic dance: 14.4 (5.0)		Ice skating: 13.9 (5.3) ² Hip-hop dance: 12.9 (4.7) ¹ Body conditioning: 13.9 (4.1) ² Aerobic dance: 11.2 (4.3) ¹		n/a	n/a	
Li, M., Fang, Q., Li, J., Zheng, X., Tao, J., Yan, X., Lin, Q., Lan, X.,	Baduanjin exercise 1hr/day 5x week x12 weeks	Self-esteem (SES Scale)	N=101 31.17 (3.69)	N=105 31.41 (3.29)	N= 96 (101 included in ITT analysis) 31.56 (3.30)	N= 105 (105 included in ITT analysis) 31.31 (3.27)	N= 93 (ITT analysis) 30.81 (3.45)	N= 101 (ITT analysis) 31.0 (3.71)	• not blinded • participants recruited from one medical university
		Mood / mindfulness (POMS scale)	N=101 102.3 (16.14)	N=105 103.5 (17.34)	N= 96 (101 included in ITT analysis) 106 (15.68)	N= 105 (105 included in ITT analysis) 107.4 (17.95)	N= 93 (ITT analysis) 103.8 (16.78)	N= 101 (ITT analysis) 104.6 (16.89)	

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
Chen, B., Zheng, G. and Chen, L. (2015)		QoL (WHOQOL-BREF)	N=101 55.84 (6.65)	N=105 54.94 (6.45)	N= 96 (101 included in ITT analysis) 55.09 (6.93)	N= 105 (105 included in ITT analysis) 54.26 (7.02)	N= 93 (ITT analysis) 56.29 (7.45)	N= 101 (ITT analysis) 55.61 (7.45)	<ul style="list-style-type: none"> • greater proportion of female participants • small effect size • <i>excellent protocol adherence</i> <i>no significant loss to follow-up</i>
		Attention (Schulte Grid)	N=101 213.9 (58.84)	N=105 224.6 (47.52)	N= 96 (101 included in ITT analysis) 192.4 (47.14)	N= 105 (105 included in ITT analysis) 210.4 (54.15)	N= 93 (ITT analysis) 193.9 (54.31)	N= 101 (ITT analysis) 202.8 (58.34)	
		Stress (CPSS Scale)	N=101 24.22 (5.18)	N=105 23.91 (5.50)	N= 96 (101 included in ITT analysis) 23.53 (5.40)	N= 105 (105 included in ITT analysis) 22.60 (5.43)	N= 93 (ITT analysis) 22.72 (5.72)	N= 101 (ITT analysis) 23.22 (5.72)	
Lindgren EC, Baigi A, Apitzsch E, Bergh H. (2010)	empowerment-based exercise intervention programme 2x week for 6 months	The Swedish version of a 10-item General Self-efficacy Scale	N= 54 Median (IQR) 32.0 (11.0-54.0)	N= 56 Median (IQR) 32.0 (14.0-47.0)	N= 27 Median (IQR) 28.0 (15.0-48.0)	N= 35 Median (IQR) 35.0 (16.0-48.00)	n/a	n/a	<ul style="list-style-type: none"> • small sample size • high dropout rate
		Social Barriers to Exercise Self-efficacy Questionnaire	N= 54 Median (IQR) Support: 9.0 (3.0-18.0) Social: 22.0 (7.0-35.0)	N= 56 Median (IQR) Support: 8.0 (3.0-16.0) Social: 18.5 (7.0-37.0)	N= 27 Median (IQR) Support: 8.0 (3.0-17.0) Social: 19.0 (7.0-36.0)	N= 35 Median (IQR) Support: 7.0 (3.0-18.0) Social: 19.0 (8.0-31.0)			

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
Noggle, J.J., Steiner, N.J., Minami, T. and Khalsa, S.B.S. (2012)	A Kripalu-based yoga program, 2-3x wk x 10 wks (28 yoga sessions total) compared to PE as usual	Profile of Mood States - Short Form	N=36 Mood Disturbance (-):42.8 (19.3) Tension anxiety (-):6.4 (4.7) Depression-dejection (-):5.1 (5.0) Anger Hostility (-):6.5 (4.7) Vigor activity (+):9.8 (4.4) Fatigue inertia (-): 8.3 (5.7) Confusion bewilderment (-): 6.8 (3.5)	N=15 Mood Disturbance (-):44.5 (10.2) Tension anxiety (-):6.7 (2.8) Depression-dejection (-):4.9 (3.0) Anger Hostility (-):6.3 (2.7) Vigor activity (+):10.2 (3.8) Fatigue inertia (-):9.8 (4.5) Confusion bewilderment (-): 6.6 (2.7)	N=35 Mood Disturbance (-):38.4 (16.9) Tension anxiety (-):5.1 (3.6) Depression-dejection (-):4.7 (4.9) Anger Hostility (-): 5.7 (5.0) Vigor activity (+):9.3 (4.0) Fatigue inertia (-): 7.2 (5.2) Confusion bewilderment (-): 6.3 (3.5)	N=15 Mood Disturbance (-):51.2 (20.1) Tension anxiety (-):9.3 (5.8) Depression-dejection (-):6.3 (4.2) Anger Hostility (-):7.1 (4.5) Vigor activity (+):10.9 (3.5) Fatigue inertia (-):9.3 (4.6) Confusion bewilderment (-): 8.3 (4.1)	n/a	n/a	<ul style="list-style-type: none"> • Small sample size. • would have been ideal to randomise individually but being in a school setting required allocation at the classroom level • Moderate attendance at the yoga classes
		Stress (Perceived Stress Scale)	N=36 19.2 (7.4)	N=15 19.1 (3.8)	N=35 18.6 (6.2)	N=15 20.3 (5.4)	n/a	n/a	

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
		Inventory of Positive Psychological Attitudes	N=36 Positive psych attributes (+):4.5 (1.0) Life Purpose/satisfaction (+):4.7 (1.0) Self conf during stress (+): 4.2 (1.0)	N=15 Positive psych attributes (+):4.5 (0.78) Life Purpose/satisfaction (+):4.8 (0.94) Self conf during stress (+): 4.2 (0.67)	N=35 Positive psych attributes (+):4.5 (1.2) Life Purpose/satisfaction (+):4.8 (1.1) Self conf during stress (+): 4.3 (0.98)	N=15 Positive psych attributes (+):4.2 (0.88) Life Purpose/satisfaction (+):4.6 (0.88) Self conf during stress (+): 4.0 (0.90)	n/a	n/a	
		Self-regulatory skills (Resilience Scale)	N=36 132.9 (18.4)	N=15 132.1 (12.4)	N=35 131.9 (24.5)	N=15 127.9 (23.4)	n/a	n/a	
		Positive and Negative Affect Schedule for Children	N=36 Positive affect (+):50.1 (11.5) Negative affect (-): 32.1 (12.5)	N=15 Positive affect (+):47.7 (9.4) Negative affect (-): 28.8 (7.7)	N=35 Positive affect (+):48.6 (11.7) Negative affect (-): 29.4 (11.5)	N=15 Positive affect (+):49.2 (11.3) Negative affect (-): 38.4 (15.5)	n/a	n/a	
		Child Acceptance Mindfulness Measure	N=36 53.9 (8.6)	N=15 52.3 (6.7)	N=35 53.4 (7.8)	N=15 49.4 (7.2)	n/a	n/a	
Staiano, A.E., Abraham, A.A. and Calvert, S.L. (2013)	Nintendo Wii Active game for 30-60 minutes per school day in a	Self-efficacy (Exercise Confidence Survey)	Cooperative (n = 19): 38.16 (12.12) Competitive (n = 19): 36.37 (13.97)	n = 16 37.38 (12.07)	Cooperative (n = 18): 42.11 (13.58) Competitive (n = 17): 37.65 (10.03)	n = 14 34.57 (11.75)	Cooperative (n = 14): 43.29 (13.40) Competitive (n = 11): 38.82 (8.82)	n = 10 35.30 (8.76)	<ul style="list-style-type: none"> Sample bias: small sample from one school and

Author (date)	Intervention	Outcome description	Baseline		Follow up 1		Follow up 2		Limitations (risk of bias)
			Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	Intervention Numbers Mean (SD)	Control Numbers Mean (SD)	
	lunch-time or after-school program. Vs. working with a peer to expend calories and earn points together	Self-esteem (Rosenberg Self-Esteem scale)	Cooperative (n = 19): 22.79 (4.45) Competitive (n = 19): 23.74 (6.47)	N=16 22.69 (3.96)	Cooperative (n = 18): 22.67 (5.91) Competitive (n = 18): 23.11 (4.78)	N=15 22.40 (5.38)	Cooperative (n = 13): 24.08 (3.88) Competitive (n = 9): 22.33 (5.74)	N=11 20.45 (5.82)	some attrition
		Peer support (Friendship Quality Questionnaire)	Cooperative (n=19): 71.89 (12.43) Competitive (n=19): 64.37 (19.58)	N=16 70.13 (18.16)	Cooperative (n = 18): 75.22 (13.39) Competitive (n = 18): 72.44 (10.78)	N=15 72.33 (17.15)	Cooperative (n = 11): 80.18 (8.59) Competitive (n = 13): 76.92 (14.04)	N=10 59.70 (20.67)	

Key

¹p<0.05 from baseline to follow up within groups, ²p>0.05 from baseline to follow up within groups

comparison with initial cycle *p<0.05, **p<0.01, ***p<0.001, comparison with 2nd cycle +p<0.05, ++p<0.01, +++p<0.001, comparison between pre- and postmenstrual phase #p<0.05, ##p<0.01, ###p<0.001

Evidence on wellbeing outcomes for sport activities and dance: synthesis of results

There are a range of wellbeing measures and interventions included in the published studies in this review. We synthesise the data in terms of intervention type and relevant wellbeing outcomes. Relevant grey literature including evaluation reports, policy documents and primary research not available in peer reviewed literature is also included.

Evidence on meditative / therapeutic physical activity and wellbeing

Published studies on meditative/therapeutic physical activity

Three RCT studies assessed the effectiveness of meditative practices including yoga (Kanojia et al., 2013; Noggle et al., 2012) and Baduanjin-Qigong (Li et al., 2015) on wellbeing in young people. The largest of these was the study by Li et al. (2015) who employed a randomized, single-blind, parallel-controlled trial to evaluate the effectiveness and safety of Baduanjin exercise (traditional Chinese body conditioning practices) for enhancing the physical and mental health of 222 Chinese medical students, aged 18-25 years and predominantly female (85.1% intervention group; 80% control group). The intervention involved the practice of Baduanjin for 1 hour a day, 5 days / week over 12 weeks with a 12 week follow up. Participants in the intervention group compared to those in the control group had no significant improvements in wellbeing related outcomes after 12 weeks, this included self-esteem as measured by Rosenberg's Self Esteem Scale ($p=0.28$), quality of life measured using the WHOQOL-BREF ($p=0.8$), self-symptom intensity (a scale covering 10 symptoms including anxiety and depression) ($p=0.71$), mood and mindfulness assessed by the Profile of Mood States questionnaire ($p=0.68$) and stress, measured by the CPSS scale ($p=0.21$). Only attention, included in the study as a primary well-being outcome, had a significant improvement with a mean Shulte Grid test score decline of 21.51 ($SD=55.7$) seconds over the intervention period comparing with a mean decline of 14.24 ($SD=60.7$) seconds in controls ($p = 0.03$), however after a 12 week follow up period there was no longer a significant difference in the Baduanjin group compared to their baseline results -20.04 [$SD=65.86$] seconds in the intervention vs -21.81 [$SD=69.14$] in the control, $p = 0.42$). The authors concluded that regular Baduanjin exercise may be an effective and safe form of exercise to promote University students' attention through a focus on adjusting breathing and mindfulness.

Two small-scale RCTs (Kanojia et al., 2013; Noggle et al., 2012) examined the effectiveness of yoga on wellbeing in young people. Kanojia et al. (2013) delivered yoga sessions for 35-40 minutes/day, 6 days/week for the duration of 3 menstrual cycles in 50 females aged 18-20 years in India. Questionnaires developed by DIPAS (Defense Institute of Physiology and

Allied Sciences), New Delhi, India were used to measure anger, trait anxiety, depression and subjective wellbeing. Results were analysed using a one-way ANOVA. In the yoga group, from the first to the 3rd cycle anxiety reduced by a mean of 3.4 points ($p<0.05$) postmenstrual and 6.16 ($p<0.001$) premenstrual, depression reduced by 3.72 points ($p<0.001$) postmenstrual and by 5.96 ($p<0.001$) premenstrual, and wellbeing scores reduced by 4.52 ($p<0.001$) postmenstrual from initial to 3rd cycle and 13.68 ($p<0.001$) premenstrual (lower scores=higher wellbeing). Anger reduced significantly in the premenstrual phase ($p<0.001$) from first to third cycle but not in the postmenstrual phase. There were no significant changes in the control group. Regular yoga practice was found to improve psychological wellbeing outcomes in young women through the menstrual cycle.

Noggle et al. (2012) examined psychological wellbeing outcomes of students attending a public high school in rural western Massachusetts (USA) in a cluster randomised study design. Participants were randomised into PE (physical education) as usual or a 10 week Kripalu-based yoga programme delivered 2-3 times a week (total 28 yoga sessions). A total of 51 participants ($n=36$ intervention; $n=15$ control) with a mean age of 17.1 years (intervention) and 17.3 years (control) was reported. A range of wellbeing measures were administered a week before and after the intervention. Analyses of covariance (ANCOVAs) were employed to assess between-group effects on end-program scores with baseline scores as the covariate. Total mood disturbance (global POMS-SF score) was significantly improved in the yoga intervention group compared with PE as usual (*effect size* = 0.689, $p=0.015$). The Tension-Anxiety subscale was also significantly improved in the yoga group (*effect size* = 0.870, $p=0.002$). A similar trend for improvement (non-significant) was found for the Confusion-Bewilderment Subscale ($p=0.053$). Other mood subscales were not significantly different between groups (Depression-Dejection, Anger-Hostility, Vigor-Activity, or Fatigue-Inertia). Negative affect (PANAS-C) was significantly improved in the yoga group (*effect size* = 0.659, $p=0.006$). Positive affect was not significantly different between groups ($p=0.54$). There were no statistically significant differences in perceived stress (Perceived Stress Scale), the Inventory of Positive Psychological Attitudes-32R (IPPA) positive psychological attitudes, self-confidence during stress, or life purpose and satisfaction subscales, nor were there significant differences in individual resilience (Resilience Scale – RS) between groups. There was a trend of greater mindfulness/acceptance in the yoga group (Child Acceptance and Mindfulness - CAMM measure) but this was not statistically significant ($p=0.097$).

No grey literature was included in the evidence on meditative/therapeutic physical activity and wellbeing.

Evidence on group sport or dance and wellbeing

Published studies on group sport or dance and wellbeing

Three published studies examined the wellbeing outcomes of group sport activities (Amorose et al., 2009; Staiano et al., 2013; Lindgren et al., 2010). Amorose et al. (2009) conducted a cohort study (pre/post-test design) to examine changes in adolescent sports player's need satisfaction (competency, autonomy, relatedness) over the course of a competitive volleyball season (in the USA) to understand how participation in sport leads to positive wellbeing outcomes (higher self-esteem and lower burnout) for young people. Staiano et al., (2013) employed an RCT design using a 20-week exergame intervention (Nintendo Wii Active Game - a sports/fitness video game requiring gross motor activity) to examine weight loss and psychosocial outcomes for African American adolescents. Lindgren et al., (2010) also used an RCT design to look at the effects of a 6-month empowerment-based exercise intervention programme on self-efficacy in inactive adolescent girls in Sweden.

Amorose et al. (2009) followed a cohort of female adolescent volleyball players (n=93; 90.6% Caucasian; mean age 15.78 years) through a season of competitive volleyball games (approximately 4 months). All were members of a large competitive club volleyball programme (n=225; 20 teams) in Midwestern U.S. Self-report questionnaires on need satisfaction (Intrinsic Motivation Inventory for sport competence; Hollembeak and Amorose autonomy scale; Richer and Vallerand's relatedness scale) and wellbeing (Rosenberg's Self-Esteem Scale; Athlete Burnout Questionnaire) were administered to participants 1-2 weeks before the start of the competitive season and 1-2 weeks after the last official game. Descriptive statistics found need satisfaction to be significantly and positively related to self-esteem and significantly and negatively related to burnout at pre- and post-season time points. Paired t-tests comparing the mean pre- and post-season scores for each variable revealed no significant changes in any variable. Considerable intra-individual variability was reported in the changes in perceived competence (*Mean Change*= 0.21, *SD* = 1.44), autonomy (*MChange*=0.03, *SD*=.94), relatedness (*MChange*=-0.05, *SD*=1.76), self-esteem (*Mchange*=0.00, *SD*=.62), and burnout (*MChange*=-0.10, *SD*=0.92). Hierarchical regression analysis to explore whether changes in need satisfaction predicted changes in wellbeing found that for self-esteem, pre-season measures did not account for significant amounts of variance in post-season self-esteem ($R^2 = .08, p > 0.05$). Adding post-season need satisfaction variables added to the prediction significantly (*Change R^2* = 0.34, $p < 0.01$, *total R^2* = 0.42). Standardised regression coefficients showed that changes in perceptions of competence ($\beta = 0.50$) and autonomy ($\beta = 0.28$) over the course of the season were positively and significantly related to changes in the athlete's self-esteem but relatedness

was not ($\beta = -0.01$). For burnout, pre-season scores did not significantly predict post-season burnout ($R^2=0.08$, $p > 0.05$). A significant effect was found after adding post-season need satisfaction ($\text{Change } R^2 = 0.39$, $p < 0.01$, $\text{total } R^2 = 0.47$). All 3 needs; perceived competence ($\beta = -0.20$), autonomy ($\beta = -0.38$), and relatedness ($\beta = -0.21$) significantly predicted burnout after controlling for pre-season scores on variables. Changes in perceptions of competence, autonomy, and relatedness over the course of the season were found to be negatively related to changes in the athletes' burnout. Consistent with Basic Needs Theory, satisfaction related positively to self-esteem and negatively to burnout. Changes in athletes' need to feel competent, autonomous and connected to others over the course of the sporting season related to changes in their overall wellbeing. Attempts to satisfy athletes' need for competence, autonomy, and relatedness by coaches, parents, and significant others may contribute to enhanced athlete wellbeing and promote more positive sport experiences.

Staiano et al., (2013) recruited 54 overweight or obese African American adolescents (15-19 years; 55.6% female; mean BMI percentile at baseline 94.7 ($SD=6.0$)) from an urban public high school to an exergame programme. BMI percentile between 85% - 95% is overweight and over 95% is obese. Participants mean BMI percentile was in the overweight category. Twenty participants dropped out due to school truancy, lack of interest, and time conflicts. Participants were randomly assigned to competitive exergame $n=19$ (compete against paired opponent for maximum points and calories burned), cooperative exergame $n=19$ (cooperate with paired opponent for maximum points and calories burned), or control conditions $n=16$ (usual daily activities e.g. socializing with friends, tutoring, sports team practice). Participants in exergame conditions played the Nintendo Wii Active Exergame 30-60 mins every school day during the lunch period or after school for up to seven months. Each gaming session consisted of pre-determined routines, increasing in difficulty over time and including cardio, upper and lower body strength training, and sports games including basketball, inline skating, baseball, tennis, and volleyball. Self-reported measures of self-efficacy (Rosenberg Self-Esteem Scale), self-esteem (Exercise Confidence Survey), and peer support (Friendship Quality Questionnaire) were taken. Growth curve modelling was used to create individual growth curve trajectories of changes in weight, self-efficacy, self-esteem, and peer support during the exergame intervention using 3 data collection points (i. baseline; ii. 10 weeks, $M=69$ days, $SD=46$; and iii. 20 weeks, $M=135$ days, $SD=48$). The growth curve for self-efficacy revealed that the cooperative condition increased in self-efficacy significantly more than the control group ($p=0.005$), but there was no difference between the competitive and cooperative groups ($p=0.172$) or competitive and control groups ($p=0.083$). The growth curve for peer support showed that the competitive condition increased significantly more in peer support than the control group ($p=0.001$). The cooperative condition was also significantly improved compared to the control ($p=0.010$), but there were no differences between the cooperative and competitive conditions ($p=0.404$). Exergaming increased peer support and self-efficacy over time and the authors conclude that cooperative exergaming may promote group cohesion and provide social

reinforcement that sustains exergame play. Cooperation in exergaming may foster a team bond more so than competition which may help overweight or obese adolescents persist during physically demanding tasks. Individual competition in game play may have been too challenging and not rewarding enough for overweight and obese adolescent participants. Negative moods associated with such competitive conditions may make them less appealing to overweight and obese youth.

Lindgren et al., (2010) carried out a stratified randomized study which included one pre- and one post-test (at six months). Eight schools were randomised (4 intervention, 4 control) and 110 non-physically active adolescent girls (13–19 years old) were assigned to an intervention group (n = 54) or a comparison group (n = 56), 43% did not complete the post tests. The intervention was a six-month empowerment-based exercise intervention programme (EIP). The EIP sessions were offered twice weekly and included 45 minutes of moderate exercise and 15 minutes of discussion, covering topics such as healthy living. Self-efficacy was measured using the Swedish version of a 10-item General Self-Efficacy Scale (GSES) and the Social Barriers to Exercise Self-efficacy Questionnaire (SPBESQ) which looks at specific self-efficacy (support barriers or social barriers). Pre- and post within group comparison showed that the intervention group had significantly improved their general perceived self-efficacy (*pre-test* $M=32.0$ ($11.0-54.0$), *post-test* $M=28.0$ ($15.0-48.0$), $p = 0.004$), whereas the control group had not improved (*pre-test* $M=32.0$ ($14.0-47.0$), *post-test* $M=35.0$ ($16.0-48.00$)). The SPBESQ scores showed no changes in the intervention or the control group. The authors conclude that EIP had a positive impact on adolescent girls' general perceived self-efficacy but not on their perceptions of their confidence to cope with barriers to participating in exercise.

Two RCT studies examined the wellbeing outcomes of group dance activities (Akandere and Demir, 2011; Kim and Kim, 2007). Akandere and Demir (2011) investigated the effect of dance on depressive feelings of a healthy student population in Turkey. 120 male and female students attending a music conservatory were recruited (20-24 years, experienced in dance activities as part of academic programme; weekly volume of physical activity = 8-10 hours). Participants were randomly allocated to the intervention dance training group (n=60; male n=30, female=30) or the control group (n=60; male=30, female=30). The dance training intervention involved Rumba and Waltz (Vals) based movements on 3 days a week for 12 weeks. Levels of depression were measured before and after the 12-week dance intervention using the Beck Depression Scale. Those receiving the dance intervention had a mean score of 15.72 ($SD=7.004$) at the start of the course, after the 12 weeks their depression score reduced significantly ($M=13.90$, $SD=5.568$, $P=0.000$). There was no reduction in depression scores in the control group (*pre-test*: $M=16.53$, $SD=5.922$; *Post-test*: $M=17.48$, $SD=7.740$, $P=0.448$). These results showed that the dance training was found to be effective in lowering self-reported feelings of depression for male and female students.

Kim and Kim (2007) examined the subjective wellbeing effects of dance, sport and exercise interventions in Korean high school and university undergraduate students. Four different types of exercise were compared in identifying the mood benefits of physical activity (continuous aerobic exercise, hip-hop freestyle dancing, ice skating for beginners involving walking and stepping on the ice, and body conditioning using weight training, jogging and stretching). 277 students (n=45 high school, n=232 university undergraduates; 17-22 years SD1.6) were randomly allocated to the four experimental groups: aerobic exercise (n = 84, 40 male, 44 female); body conditioning (n = 64, 44 male, 20 female); hip-hop dancing (n =45, 45 female) and ice skating (n = 84, 60 male, 24 female). Each group took part in physical activity for a one-off session of 40 minutes. Mood changes (positive wellbeing, psychological distress and fatigue) were measured before and after the exercise sessions based on a Korean translation of the Subjective Exercise Experiences Scale. A 4 Groups x 2 Tests (pre- and post-test) analysis of variance with repeated measures was conducted to examine the effects of exercise on mood states. Results showed that aerobic dance significantly increased all three wellbeing outcomes (*positive wellbeing: M=16.8 (SD=4.0) pre-test, M=19.9 (SD=3.9) post-test, P<0.05; psychological distress: M=9.4 (SD=4.2) pre-test, M=6.7 (SD=2.9) post-test, P<0.05; fatigue: M=14.4 (SD=5.0) pre-test, M=11.2 (SD=4.3) post-test, P<0.05*) as did hip-hop dance (*positive wellbeing: M=16.3 (SD=4.2) pre-test, M=19.7 (SD=3.4) post-test, P<0.05; psychological distress: M=9.8 (SD=4.6) pre-test, M=7.3 (SD=4.2) post-test, P<0.05; fatigue: M=16.2 (SD=4.4) pre-test, M=12.9 (SD=4.7) post-test, P<0.05*). Taking part in body conditioning and ice skating resulted in no significant improvements for positive wellbeing, psychological distress or fatigue. Overall, mood alteration appeared to vary with the activity but the findings indicate that a brief exercise program leads to reported positive well-being and less psychological distress.

Grey literature on group sport or dance and wellbeing

Three evaluation reports presented recent findings on a variety of community-based group sport or dance activities and wellbeing and included sport (Mansfield et al., 2016), dance (Potter et al., 2015), and circus skills (BOP consulting, 2016).

Mansfield et al., (2016) reported on the qualitative findings of the Health and Sport Engagement (HASE) Project (Mansfield et al., 2015) which aimed to engage previously inactive people in sustained sporting activity for 1 x 30 minutes a week and examine the associated health and wellbeing outcomes of doing so. Between September 2013 and September 2014, 37 HASE partners (8 lead partners, 29 associate partners) were involved in delivering and evaluating 15 different HASE delivery projects involving more than 24 HASE sports, to 417 participants representing 11 distinct inactive groups in the London Borough of Hounslow (UK). The longitudinal process evaluation employed focus groups (n=26) with potential participants (n=220) prior to the intervention to understand issues of inactivity and support the design of community sports, interviews with delivery personnel (n=30) to

understand the design and delivery elements of the project, and interviews with participants (n=15) alongside structured observations of community sport projects to understand the experiences of inactive people in community sport. Young participants (14-24 years) took part in school and community-based sport activities including martial arts, dance, gym-based exercise, exercise classes, swimming, netball, cycling and football. Findings illustrated there are negative and positive aspects of wellbeing associated with engagement in community sport. People who are currently inactive can hold deep-rooted negative feelings about taking part in sport which contributes to the perception that sport is not an activity that can enhance wellbeing. Concerns relate to personal capability and competence and unfavourable comparisons to those who are 'sporty'. Different preoccupations are evident at different life phases and adjusting to change is particularly significant in perceptions of sport participation and wellbeing. Young people are especially self-conscious about comparisons with peers yet they also hold personal perceptions of the positive contribution that physical activity can have on health and wellbeing. Appropriately tailored sporting activities (low intensity, accessible, low or no cost, short duration, varied, informal, appropriate and interesting) can reinforce more positive perceptions about wellbeing and physical activity, encourage participation and enhance wellbeing outcomes from sport engagement including social connectedness, feelings of pleasure and the development of a sense of purpose.

Potter et al., (2015) report on the evaluation of the DanceQuest (2012-2015) project, a national UK programme aiming to encourage young people (11-13 years) to perform and watch dance, engage with local dance venues, spaces and dancers, and provide key stage 3 (KS3) teachers in training for using dance in the curriculum. Across the 3 years (2012-2015) the project involved 19 participating schools, 25 KS3 teachers, 12 project team members, 18 dance/music practitioners and the delivery of 353 in-school workshops to 1498 young people. 10 training events were delivered to 84 teachers. 24 live performances were watched by 3079 young people and 350 adults. 12 events were performed by 2096 young people and watched by 2996 adults. Evaluation methods included a short evaluation questionnaire to participant students (on completion of the DanceQuest programme) and teachers (at baseline and completion) and interviews, observations, and photographs employed throughout the project. Key findings from the questionnaires illustrated that dance had made a positive impact on physical health and wellbeing (90%), a positive impact on mental health and wellbeing (86%) and a positive impact on student's future aspirations (76%). Findings from the qualitative data showed that dance performance instilled wellbeing feelings associated with exhilaration and sense of purpose, and increased confidence and self-esteem.

BOP Consulting (2016) conducted an evaluation of the social impact of Jacksons Lane's community outreach programme in the London (UK) Boroughs of Tottenham and Haringey. Jacksons Lane is a multi-arts venue based offering a programme of contemporary circus,

comedy, dance and performance. For children and young people (8-21 years), Jackson Lane's programmes include circus, music, drama, clowning and, heritage workshops, run by professionals and artists and supported by volunteers. Additionally, there are opportunities for young people to develop careers in the arts services, talent and skills development services and volunteering as well as personal development opportunities. The evaluation was designed around reviewing existing literature, evaluation and monitoring data, attendance and observation at 13 sessions and events, 56 short semi-structured interviewees with participants, including 23 young people (8-21 years), and short semi-structured interviews with volunteers and staff. For young people on circus and leadership projects impact was strongest and most widespread in terms of increased confidence, improved interpersonal skills, (collaboration, communication), increased wellbeing (happiness, relaxation, playfulness, fun), gaining creative skills and expression, increased social bonds, combatting isolation, improved sense of being a part of something positive, and increased aspiration and ambition. Longer term impacts were identified in the evaluation in relation to educational attainment and workforce readiness, social engagement and sense of belonging.

Discussion

Summary of key findings

There is a small amount of high quality evidence from 2 published RCT studies that therapeutic-type physical activity (yoga and Baduanjin Qigong) can improve feelings of anxiety, depression, anger, attention and overall subjective wellbeing. There is moderate quality evidence from 1 RCT study for significant effects of yoga-type activity on wellbeing in terms of improving total mood disturbance, tension-anxiety and negative affect but not on perceived stress, positive psychological attitudes, self-confidence, life purpose and satisfaction, positive affect, resilience, or mindfulness, or on scales relating to confusion-bewilderment, depression-dejection, anger-hostility, vigor-activity and fatigue-inertia.

In terms of group sport and dance there is low quality evidence (1 study) that for players taking part in a competitive club environment (volleyball), improved feelings of competency, positive relationships with others, and a sense of autonomy within the group may lead to improved player wellbeing and promote a positive sport experience. There is low-moderate quality evidence (1 study) that cooperative (peer-support) exergaming may promote group cohesion and provide social reinforcement that sustains exergame play in overweight African-American young girls. Additionally, there is moderate quality evidence (1 study) that exercise programmes focusing on empowering young girls to take part can have a positive impact on general perceived self-efficacy but not on perceptions of confidence to cope with barriers to participating in exercise. Moderate quality evidence (1 study) is provided for significant positive mood enhancement through taking part in aerobic dance and hip-hop

dance compared to ice skating and body conditioning. There is moderate quality evidence (1 study) to show that dance training can be effective in lowering self-reported feelings of depression for male and female students.

Included grey literature focused on the impact of group sport and dance interventions on young healthy people. Findings from qualitative research about community sport and wellbeing (1 report) illustrate negative and positive aspects of wellbeing associated with taking part. Negative feelings about sport participation amongst young people, including concerns about competency and capability can create the view that sport and physical activity are not wellbeing-related. At the same time young people also recognise the potential for a more positive contribution of physical activity to health and wellbeing. Appropriately tailored sport and physical activities can enable young people to discover positive experiences and feelings from taking part particularly in relation to social connectedness, feelings of pleasure and the development of a sense of purpose. There is evidence of wellbeing enhancement in taking part in community dance (1 study). Qualitative findings showed that dance performance (both taking part and watching) instilled wellbeing feelings such as exhilaration and sense of purpose, and increased confidence and self-esteem. Evidence from a community outreach project (1 study) that employed a circus skill programme in engaging young people demonstrated impact in terms of increased confidence, interpersonal skills, (collaboration, communication), wellbeing (happiness, relaxation, playfulness, fun), creative skills and expression, social bonds, sense of being a part of something positive, aspiration and ambition and reducing isolation.

The evidence from the published studies overall is limited, very selective, drawn from very varied national and cultural contexts, and mostly of moderate-low quality. A wide variety of wellbeing measures were used and there was a great deal of heterogeneity across the studies. Meta-analysis was not appropriate in this systematic review and the match between the wellbeing measures used in the published studies with the ONS4 subjective questions is very imprecise. The findings from the eligible published studies provide very limited, specific evidence of significant wellbeing outcomes associated with exercise interventions delivered for the purposes of the research, and to particular participants. It is not clear whether the findings from these studies are generalizable to other sport or dance settings or apply in everyday life. The scope of the eligible studies is selective in terms of countries and types of intervention. Most of the interventions are in educational settings, with school or college students. There is a big gap in terms of understanding the wellbeing impacts of community-based, recreational, team, and both formal and informal sport and dance participation; there is no representation of the forms of sport with the highest participation in the UK (swimming, athletics, cycling).

There is a very limited focus on understanding the precise context and mechanisms of intervention effectiveness in the included published studies. A limited number of low quality studies claimed the group dynamic aspect of sport and dance to be significant in enhancing wellbeing outcomes. It is not possible to conclude from this whether formally organised, club-based delivery mechanisms are more efficacious in enhancing wellbeing than informal delivery models.

The number of eligible grey literature studies was small (n=3). Data from surveys was small-scale which may limit the strength of the findings and generalisability of the evidence. However, the high level of detail, analytical rigor and a focus on context specific processes and outcomes means the findings of the strongest reports are credible.

Completeness of the included evidence

The review includes 7 randomised control trials and 1 cohort study (pre/post-test design) from the published literature. There are few published papers examining the subjective wellbeing outcomes of sport and dance for healthy young people (aged 15-24 years) and none of these considered UK populations. No published qualitative studies met our inclusion criteria yet high quality qualitative research methods have the potential to reveal rich insights into the wellbeing benefits of sport and dance. Such insights are revealed in the higher quality grey literature in this evidence review which is UK focused.

Quality of the included evidence

The review includes 2 high quality RCT studies, 4 moderate quality RCT studies, 1 low quality RCT and 1 low quality cohort study. In terms of quality assessed via the GRADE approach there appears to be better quality evidence in the studies on yoga-type activity using controlled designs but these have yielded mixed results. Despite the 2 high quality RCTs, overall the quality of evidence on the wellbeing outcomes of sport and dance for health young people (15-24 years) is moderate-low in respect of there being very little evidence in total, the (small) sample sizes (range n=222 to n=50) and sample bias.

The evidence from the grey literature was produced through systematic qualitative and quantitative methods. The strongest reports included a high level of descriptive and theoretical detail about evaluation methods and acknowledged the limitations of evaluation design. It was not always clear how themes were identified and developed and it was not always apparent that conclusions emerged from comprehensive data treatment. One report made a clear attempt to search for disconfirming cases and consider the negative wellbeing impact of sport participation (see Mansfield et al., 2016) but evaluation reports tended to focus only on the positive impacts of sport and dance. Further, there was a tendency in

some evaluations to rely on face value reporting of participants' accounts rather than developing latent forms of thematic analysis where appropriate (Braun & Clarke, 2006).

Most published studies obtained appropriate ethics approval, although this was not always reported. Few studies provided exact details of the researcher's role, potential bias and influence on sample recruitment, setting and responses of participants.

Strengths and Limitations of the review process

The large number of hits following initial searches and the overlap between elite sport performance and measures of state and trait anxiety means that it is possible that some relevant evidence has not been included in this report. The focus on a specific target age group will have excluded evidence from studies that have aggregated data across younger and older age groups in their analysis. However, the comprehensive search strategy ensures that this overview represents a comprehensive summary of all existing eligible studies published prior to the search dates and the pre-publication of our protocol on PROSPERO ensures methodological transparency and militates against potential post-hoc decision making which can introduce bias to the process. Dual screening of searches and data extraction and independent quality assessment of included reviews ensured a rigorous process.

Taking published studies as the sole evidence increases the potential risk of publication lag, wherein possible important new evidence that has not yet been included in published reports is not identified and included.

The use of the GRADE criteria introduces an element of subjective judgement. A consistent approach to judgements across the different interventions has been applied but it should be recognised that these judgements are open to interpretation.

Implications for research, policy and practice

No UK studies were eligible for inclusion in this review. It is not possible to conclude that findings in this review are generalizable across countries. There is a need for UK national and local policy in sport and dance to recognise the potential significance of participation on subjective wellbeing and to support programmes that focus on building the evidence.

Recent national sport strategy in the UK (DCMS, 2015; Sport England, 2016) identifies wellbeing as an outcome for sport and physical activity and needs to be accompanied by attention to agreeing definitions and developing relevant measures of wellbeing outcomes, and evaluating what works to enhance wellbeing in sport and dance. National agencies

across the sport, culture and health sectors (e.g. DCMS, ACE, Sport England, PHE) may be influential in promoting this approach; conversely, a lack of national lead may discourage regional and local stakeholders from prioritising this.

The lack of evidence identified in this review does not necessarily mean that wellbeing benefits are not accrued from taking part in sport and dance. There is scope to build evidence on wellbeing outcomes of sport and dance in healthy young people through well-designed, rigorous and appropriate research methods which are underpinned by relevant theory, use established methods of analysis, and focus both on measuring wellbeing outcomes and understanding the precise context and mechanisms of intervention effectiveness.

The development of a programme of wellbeing evaluation training would support key personnel in the sport and dance sectors in ensuring a comprehensive programme of delivery includes appropriate and rigorous monitoring and evaluation.

There is a need for studies of the wellbeing impacts of sport and dance for young people to be made public through academic and non-academic dissemination.

References

Included Studies: Sport, Dance and Wellbeing Review

- Akandere, M. and Demir, B., 2011. The effect of dance over depression. *Collegium antropologicum*, 35(3), pp.651-656.
- Amorose, A.J., Anderson-Butcher, D. and Cooper, J., 2009. Predicting changes in athletes' well being from changes in need satisfaction over the course of a competitive season. *Research Quarterly for Exercise and Sport*, 80(2), pp.386-392.
- Kanojia, S., Sharma, V.K., Gandhi, A., Kapoor, R., Kukreja, A. and Subramanian, S.K., 2013. Effect of yoga on autonomic functions and psychological status during both phases of menstrual cycle in young healthy females. *J Clin Diagn Res*, 7(10), pp.2133-9.
- Kim, S. and Kim, J., 2007. Mood after various brief exercise and sport modes: aerobics, hip-hop dancing, ice skating, and body conditioning. *Perceptual and motor skills*, 104(3 suppl), pp.1265-1270.
- Li, M., Fang, Q., Li, J., Zheng, X., Tao, J., Yan, X., Lin, Q., Lan, X., Chen, B., Zheng, G. and Chen, L., 2015. The effect of Chinese traditional Exercise-Baduanjin on physical and psychological well-being of college students: A randomized controlled trial. *PloS one*, 10(7), p.e0130544.
- Lindgren, E.C., Baigi, A., Apitzsch, E. and Bergh, H., 2010. Impact of a six-month empowerment-based exercise intervention programme in non-physically active adolescent Swedish girls. *Health Education Journal*, p.0017896910379366.
- Noggle, J.J., Steiner, N.J., Minami, T. and Khalsa, S.B.S., 2012. Benefits of yoga for psychosocial well-being in a US high school curriculum: a preliminary randomized controlled trial. *Journal of Developmental & Behavioral Pediatrics*, 33(3), pp.193-201.
- Staiano, A.E., Abraham, A.A. and Calvert, S.L., 2013. Adolescent exergame play for weight loss and psychosocial improvement: a controlled physical activity intervention. *Obesity*, 21(3), pp.598-601.

Additional References

- Benzies, K.M., Premji, S., Hayden, K.A. and Serrett, K., 2006. State-of-the-evidence reviews: advantages and challenges of including grey literature. *Worldviews on Evidence-Based Nursing*, 3(2), pp.55-61.
- Biddle, S.J. and Asare, M., 2011. Physical activity and mental health in children and adolescents: a review of reviews. *British journal of sports medicine*, p.bjsports90185.
- Bradt, J., Goodill, S.W. and Dileo, C., 2011. Dance/movement therapy for improving psychological and physical outcomes in cancer patients. *The Cochrane Library*.
- Cavill, N., Richardson, D., and Foster, C. (2012). *Improving Health Through Participation in Sport: a review of research and practice*. BHF Health Promotion Research Groups. Oxford University and Sport England.

- Department of Culture, Media and Sport (2015) *Sport Future: a new strategy for an active nation*. London: Crown Copyright.
- Daykin with Joss (2016) Public Health England Arts and Health Evaluation Framework https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/496230/PHE_Arts_and_Health_Evaluation_FINAL.pdf
- Dolan, P. and Galizzi, M.M., 2015. Like ripples on a pond: behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47, pp.1-16.
- Foster, C., Hillsdon, M., Thorogood, M., Kaur, A. and Wedatilake, T., 2005. Interventions for promoting physical activity. *The Cochrane Library*.
- Fujiwara, D., Kudrna, L. and Dolan, P., 2014. Quantifying and valuing the wellbeing impacts of culture and sport. *Department for Culture Media and Sport Research Paper*.
- Mammen, G. and Faulkner, G., 2013. Physical activity and the prevention of depression: a systematic review of prospective studies. *American journal of preventive medicine*, 45(5), pp.649-657.
- Mansfield, L., Anokye, N., Fox-Rushby, J. and Kay, T., 2015. The Health and Sport Engagement (HASE) Intervention and Evaluation Project: protocol for the design, outcome, process and economic evaluation of a complex community sport intervention to increase levels of physical activity. *BMJ open*, 5(10), p.e009276.
- Meekums, B., Karkou, V. and Nelson, E.A., 2015. Dance movement therapy for depression. *The Cochrane Library*.
- Ogilvie, D., Foster, C.E., Rothnie, H., Cavill, N., Hamilton, V., Fitzsimons, C.F., Mutrie, N. and Scottish Physical Activity Research Collaboration, 2007. Interventions to promote walking: systematic review. *Bmj*.
- Ogilvie, D., Egan, M., Hamilton, V. and Petticrew, M., 2004. Promoting walking and cycling as an alternative to using cars: systematic review. *Bmj*, 329(7469), p.763
- Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B. and Kohlberger, T., 2011. Health benefits of cycling: a systematic review. *Scandinavian journal of medicine & science in sports*, 21(4), pp.496-509.
- Oja, P., Titze, S., Kokko, S., Kujala, U.M., Heinonen, A., Kelly, P., Koski, P. and Foster, C., 2015. Health benefits of different sport disciplines for adults: systematic review of observational and intervention studies with meta-analysis. *British journal of sports medicine*, bjsports-2014.
- OneDanceUk (2017) *About Us*. <http://www.onedanceuk.org/>
- O'Mara-Eves, A., Brunton, G., McDaid, G., Oliver, S., Kavanagh, J., Jamal, F., Matosevic, T., Harden, A. and Thomas, J., 2013. Community engagement to reduce inequalities in health: a systematic review, meta-analysis and economic analysis. *Public Health Research*, 1(4).
- Priest, N., Armstrong, R., Doyle, J. and Waters, E., 2008. Policy interventions implemented through sporting organisations for promoting healthy behaviour change. *The Cochrane Library*.
- Ren, J. and Xia, J., 2013. Dance therapy for schizophrenia. *The Cochrane Library*.
- Start Active Stay Active, 2011 A report on physical activity from the four home countries' Chief Medical Officers. *Start Active, Stay Active: A Report on Physical Activity from the Four Home Countries' Chief Medical Officers*.

- Ströhle, A., 2009. Physical activity, exercise, depression and anxiety disorders. *Journal of neural transmission*, 116(6), pp.777-784.
- South, J., 2015. A guide to community-centred approaches for health and wellbeing. Sport England (2016). *Towards an Active Nation; strategy 2016-2021*. London: DCMS
- Wheaton, B., Gilchrist, P., Doidge, M., Caudwell, J., Jones, I., Mansfield, I., et al (2015-2018) ESRC Seminar Series Understanding the social benefits of informal and lifestyle sports ES/M002705/1. A partnership led by Brighton University with Waikato University (New Zealand), Bournemouth University and Brunel University London.
- Wheaton, B., and Gilchrist, P., (2017 forthcoming) Special Issue of the International Journal of Sport Policy and Politics: *Understanding the social benefits of informal and lifestyle sports*
- World Health Organization, 2010. *Global recommendations on Physical Activity for health*. World Health Organization.

Appendix 1: Summary of wellbeing measures (published studies)

Measurement tool	Outcome measuring	Description	Scoring/ interpretation	Validity & Reliability
<u>Rosenberg's Self-Esteem Scale</u>	Self-esteem	10-item scale that measures global self-worth by measuring both positive and negative feelings about the self. The scale is believed to be uni-dimensional. All items are answered using a 4-point Likert scale format ranging from strongly agree to strongly disagree. Five of the items have positively worded statements and five have negatively worded ones. The scale measures state self-esteem by asking the respondents to reflect on their current feelings.	Range: 0-30 15- 25 normal range; below 15 low self-esteem.	The original sample for which the scale was developed in the 1960s consisted of 5,024 high school juniors and seniors from 10 randomly selected schools in New York State and was scored as a Guttman scale. The scale generally has high reliability: test-retest correlations are typically in the range of .82 to .88, and Cronbach's alpha for various samples are in the range of .77 to .88.
Athlete Burnout Questionnaire (Raedeke & Smith 2001)	Athletes level of Burn out	15 item assessing 3 dimensions of burnout: -Emotional/physical exhaustion -Reduced sense of accomplishment -Sport devaluation The stem for each item is "How often do you feel this way?" Each response is scored on a 5-point Likert scale: "almost never" (1), "rarely" (2), "sometimes" (3), "frequently" (4), "almost always" (5).	Combined scores from each item for a single global indicator (higher the score the higher the level of burnout)	Raedeke and Smith (2001) and Cresswell and Eklund (2006) demonstrated reliability and validity both in and out of North America
<u>Beck Depression Inventory</u> first published in 1961, revised in 1978 (BDI-1A) and then 1996 (BDI-II)	Depression (presence and degree. NOT a diagnostic instrument)	21-question multiple-choice self-report inventory for adolescents and adults. Evaluates 21 symptoms of depression (15 on emotions, 4 on behavioural changes, 6 on somatic symptoms). The 21 items cover sadness, pessimism, past failure, self-dislike, self-criticism, suicidal thoughts or wishes, crying,	0–9 not depressed 10–18 mild-moderate depression 19–29 moderate-severe 30–63 severe According to paper: 0–9 normal 10–15 low	Beck reviewed 11 studies and the BDI was capable of discriminating between groups that contrasted in level of depression. Beck's original paper reported an internal consistency studies demonstrated a correlation coefficient of .86 for the test items, and the Spearman-

		agitation, loss of interest, indecisiveness, worthlessness, loss of energy, changes in sleeping patterns, irritability, changes in appetite, difficulty concentrating, tiredness or fatigue, and loss of interest in sex. Time to Administer: 5-10 minutes	16–23 medium 24+ depressive	Brown correlation for the reliability of the BDI yielded a coefficient of .93. Criticisms; BDI-IA only addresses six out of the nine DSM-III criteria for depression, self-reported (reporting bias), questionnaire therefore the way administered could affect outcome e.g. social desirability. If pt has a physical illness the physical symptoms such as fatigue may score higher but not reflect depression.
Subjective Exercise Experiences Scale (SEES)	Measuring 3 dimensions; positive well-being, psychological distress, and fatigue	“By circling a number on the scale below each of the following items, please indicate the degree to which you are experiencing each feeling now, at this point in time, after exercising”. Each item rated on a 7-point Likert scale: 1 (Not at all) - 7 (Very much so). 12 item scale (4 items per dimension): great, awful, drained, positive, crummy, exhausted, strong, discouraged, fatigued, terrific, miserable, and tired.	The Items (4 items per dimension) are summed to create a summary score for Positive Well-Being, Psychological Distress and Fatigue. Therefore each dimension has a possible score up to 28, the higher the number the higher the association with the trait	validity and reliability have been reported for other groups (McAuley & Courneya, 1994; Rudolph & Kim, 1996).
Positive Affect Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988)	Hedonic Well-Being/ the intensity associated with both positive and negative dimensions of global affect	20-item self-report instrument. Rate each using a 5-point Likert scale ranging from 1 (Not At All or Very Slightly) to 5 (Very Much). PANAS for Children (PANAS-C): 30-item measure (15 positive affect and 15 negative affect items). Indicate how often they have felt interested, sad, and so on during the “past few weeks” on a 5-point Likert scale ranging from 1 (very slightly or not at all) to 5 (extremely).	Positive Affect Score: range from 10 – 50, with higher scores representing higher levels of positive affect. Negative Affect Score: ranges from 10 – 50, with lower scores representing lower levels of negative affect. PANAS-C: Summation scores for positive affect and negative affect range from 10 to 75 each.	Reliability and Validity reported by Watson (1988) was moderately good. For the Positive Affect Scale, the Cronbach alpha coefficient was 0.86 to 0.90; for the Negative Affect Scale, 0.84 to 0.87. Over a 8-week time period, the test-retest correlations were 0.47-0.68 for the PA and 0.39-0.71 for the NA. The PANAS has strong reported validity with such measures as general distress and dysfunction,

				depression, and state anxiety. PANAS-C has demonstrated good convergent and discriminant validity in adolescent samples
Self-Esteem Scale	Self-esteem	10 items, and the total score ranges from 10 to 40.	Higher scores = higher self-esteem	NR
Profile of Mood States POMS scale	Mood & mindfulness	7 subscales (tension, anger, fatigue, depression, vigor, confusion, and mood related to self-esteem) with 40 adjectives that describe mood. Original: 65 adjectives rated on 5-point scale 0= not at all; 1=a little; 2=moderately; 3=quite a bit; 4=extremely Short Form (POMS-SF): 30-item consisting of 30 adjectives rated on a 5-point scale ranging from 0 (not at all) to 4(extremely).	Higher scores (POMS Total Mood Disturbance (TMD)) = more negative current mood states POMS-SF: Responses are summed (with positive items reverse scored) to provide a TMD score (range 0–100), as well as subscale scores for 6 mood states (each ranging 0–20): Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor-Activity, Fatigue-Inertia, and Confusion-Bewilderment.	High internal consistency of subscales and validity for original POMS scale
<u>WHOQOL-BREF Scale</u>	QoL	The World Health Organization Quality of Life (WHOQOL). WHOQOL-BREF is a shorter version containing 26 items (1 from each of the 24 facets in the WHOQOL-100 plus 2 items from the Overall quality of Life and General Health facet) measuring these domains: physical health, psychological health, social relationships, and environment.	QoL profile with 4 domain scores plus overall perception of QoL and overall perception of health. Higher scores = higher QoL.	developed by the WHOQOL Group with fifteen international field centres, simultaneously, in an attempt to develop a quality of life assessment that would be applicable cross-culturally
Schulte Grid	Attention	A Schulte table (8*8 grid) is a square that consists of 64 squares of the same size (1 × 1 cm), with one of 64 random numbers from 1 to 64. When tested, individuals are required to figure out the numbers in the order from 1 to 64, and read out the	Less time represents higher level of attention	

		numbers loud at the same time. Timing starts with 1 and ends with 64.		
<u>Perceived stress scale (PSS)</u> (Cohen et al, 1983). Chinese Perceived Stress Scale (CPSS)	Stress	10-items measuring the degree to which events are appraised as stressful during the past month. Items rated on a Likert scale from 0 (never) to 4 (very often). Items designed to tap how unpredictable, uncontrollable, and over-loaded respondents find their lives. The scale also includes several direct queries about current levels of experienced stress. CPSS-14 questions	Responses summed to give a total score ranging from 0 to 40 (CPSS: 0-56). Higher composite scores indicate greater perceived stress.	The PSS is the most widely used psychological instrument for measuring the perception of stress. CPSS-10 showed a stable two-factor structure with satisfactory internal consistency and construct validity (Siu-man Ng, 2013)
<u>Inventory of Positive Psychological Attitudes</u>	Positive worldview, Confidence in Life and Self (two sub-scales: Life Purpose and Satisfaction (LPS) and Self-Confidence During Stress (SCDS)).	32-item, 7-point Likert self-report scale. Example questions: Life Purpose and Satisfaction Section: My daily activities are - Response: not a source of satisfaction to a source of satisfaction (7 pt scale) Self-Confidence During Stress Section: When there is a great deal of pressure being placed on me - Response: I get tense to I remain calm (7 pt scale).	Each score is calculated as a mean; possible scores ranging from 1 to 7. VERY LOW: 1.00 TO 2.49 MEDIUM LOW: 2.50 TO 3.99 MEDIUM HIGH: 4.00 TO 5.49 VERY HIGH: 5.50 TO 7.00	It has been shown to possess adequate reliability and construct validity in samples of undergraduate college students.
Resilience Scale	self-regulatory skills (degree of individual resilience)	25-item covering 5 factors of resilience; meaningful life (purpose); perseverance; self-reliance; equanimity; and coming home to yourself (existential aloneness). Items scored on a 7-point scale from 1 (disagree) to 7 (agree)	Possible scores ranging from 25 to 175. Higher scores reflect greater resilience.	The scale has internal consistency, reliability, and concurrent validity and has been recommended as the best instrument for measuring resilience in adolescents (Ahern et al, 2006).
Child Acceptance Mindfulness Measure	self-regulatory skills (mindfulness)	25-item measure assessing the degree to which children and adolescents observe internal experiences, act with awareness, and accept internal experiences without judging them.	A total score is calculated by reverse scoring negatively worded items and summing the item total. Range in scores from 0 to 100. Higher scores indicate higher levels of acceptance and mindfulness.	The CAMM has demonstrated good internal consistency and concurrent validity with negative correlations to measures of cognitive suppression and

				psychological inflexibility in a study of 606 middle school students (Coyne, Cheron & Ehrenreich, 2008)
<u>State Trait Anger Expression Inventory-2TM</u>	self-regulatory skills (experience, expression, and control of anger)	<p>Designed for people aged 16 years and older.</p> <p>57-item self-report tool with a 4-point Likert response format. The instrument is categorized into subscales that reflect state anger (3 subscales), trait anger (2 subscales), and anger expression ().</p> <p>Study reported in used sub scales measuring anger expression. Anger expression was conceptualized as having 3 major components: anger-out (outward expression of anger), anger-in (anger suppression), and anger control (attempts to control expression of anger)</p>	<p>For each scale, summation scores range from 8 to 32.</p> <p>Higher the score = stronger association. Higher Anger-in = more negative anger expression, higher anger-out = more negative anger expression, higher anger control = better anger control.</p>	
Friendship Quality Questionnaire	Peer support	assess the quality of children's and early adolescents' relationships with their best friends according to five dimensions: companionship, conflict, help/aid, security and closeness.	http://spr.sagepub.com/content/11/3/471.abstract	<p>A confirmatory factor analysis, used to evaluate the factor structure of this instrument, demonstrated that these scales represented distinct, but related, domains of friendship.</p> <p>Assessments of reliability indicated the high level of internal consistency within each dimension. The validity of the scale was indicated by the observation of higher ratings for (a) mutual friends than for non-mutual friends, and (b) for stable friends than for non-stable friends.</p>
<u>Exercise Confidence</u>	Self-efficacy	8 items each on a 10 pt likert scale (I know I can, to 10 I know I cannot)	Total the numbers circled and the higher the score, the less likely you are to stick with your	

<u>Survey</u>			exercise program.	
-5 item Subscale of the Intrinsic Motivation Inventory	Need Satisfaction (sport competence)	5 items – rated on 7 pt Likert scale (1 strongly disagree – 7 strongly agree)	Higher score = agree more	Each has shown adequate psychometric properties with adolescent athletes in similar studies testing SDT (Amorose & Anderson-Butcher 2007)
-6 item Scale (developed by Hollembeak & Amorose, 2005)	Need Satisfaction (need for autonomy)	the measure asks respondents to indicate the amount of choice or control they have when participating in their current sport. 6 items - rated on a range from 1 (not at all true) to 5 (completely true)		
-Sport oriented version of Richer & Vallerand's Feelings of Relatedness Scale	Need Satisfaction (need for relatedness)	rate the extent they agree with a series of 10 adjectives describing their relationships with members of their sport team. Range from 1 (do not agree at all) to 7 (very strongly agree)		
-Anger (16 item questionnaire)	Anger	Questionnaires were developed by DIPAS (Defense Institute of Physiology and Allied Sciences), New Delhi, India. Every item amongst all the questionnaires measures the tested domain on the weighted scores of responses from 0 (never) to 3 (almost always).	Questionnaires were scored by adding the weighted (0 to 3) scores of each item.	These questionnaires were chosen as they are valid for Indian population, reliable and specific to measure the tested psychological domains.
-Trait anxiety (40 item questionnaire)	Anxiety			
-Depression (10 item questionnaire)	Depression (incl. depressed mood, guilt, difficulty in sleeping, decision making, work and interests)			
Subjective well-being (50 item questionnaire)	Subjective WB (develop potential; work			
			Sense of well-being: The lesser the score the better is the sense of well-being	

	productivity and creativity; build strong and positive relationships with others)			
The Swedish version of a 10-item General Self-efficacy Scale (<u>GSES</u>)	Self-efficacy	The Swedish version of the 10-item General Self-Efficacy Scale (GSES), developed by Koskinen-Hagman, Schwartz and Jerusalem. Original version used a 4-point Likert scale, but a pilot test demonstrated that was too limited to detect variations in participants' responses. This scale was extended to a 6-point Likert scale	The total score is calculated by finding the sum of all items. The total score ranges between 10 and 60, with a higher score indicating more self-efficacy. This paper argues that lower score indicates a higher perceived GSE. (note original scale says higher score = higher S-E)	GSES is correlated to emotion, optimism, work satisfaction. Negative coefficients for depression, stress, health complaints, burnout, and anxiety. Internal reliability for GSE = Cronbach's alphas between .76 and .90
Swedish version - Social Barriers to Exercise Self-efficacy Questionnaire (SPBESQ)	Identify behaviour changes	6-point Likert scale. Response ranges from 1 'not true' to 6 'absolutely true'. 10 items (3 for support barriers and 7 for social barriers) Examines content of intervention and specific behavioural changes.	A lower score appears to suggest a higher perceived SSBES in this paper (note lack of clarity re: direction of effect)	NR

Appendix 2: Reasons for exclusions and table of excluded studies

Reasons for Exclusion:

- **Population** - Does not include the population of interest i.e. adult participants, worldwide, healthy or unhealthy, excluding paid professionals
- **Outcome** - Does not include outcomes of interest i.e. subjective wellbeing measured as an outcome measure using a recognised measure/method
- **Intervention** - Does not include interventions of interest i.e. interventions focused on music or singing including listening, performing and music therapy offered to enhance wellbeing (Excluding clinical music therapy, clinical procedures, medical tests and diagnostics)
- **Study design** – Is not a study design of interest i.e. primary study with empirical data of wellbeing outcomes and processes by which wellbeing outcomes are achieved. Quantitative, qualitative or mixed methods. Published between 1996-2016
- **Comparator** – does not use a comparator e.g. no music or signing, white noise, usual routine i.e. inactive comparator

Authors	Year	Reason for exclusion
Adie JW, Duda JL, Ntoumanis N	2008	Intervention
Ahola R., Pyky R., Jämsä T., Mäntysaari M., Koskimäki H., Ikäheimo T.M., Huotari M.-L., Röning J., Heikkinen H.I., Korpelainen R.	2013	Study design
Altintas A., Asci FH., Kin-Isler A., Guven-Karahan B., Kelecek S., Ozkan A., Yilmaz A.,Kara FM	2014	Population
Anamaria, Constantinescu	2013	Outcome
Aphamis G., Giannaki C.D., Tsouloupas C.N., Ioannou Y., Hadjicharalambous M.	2015	Outcome
Aramendi Jauregui P., Bujan Vidales K., Arburua Goyeneche R.	2014	Intervention
Barton, J. and Pretty, J.	2010	Study Design
Berntsson, LT., Ringsberg, KC	2014	Intervention
Booker C.L., Skew A.J., Kelly Y.J., Sacker A.	2015	Intervention
Booker C.L., Skew A.J., Sacker A., Kelly Y.J.	2014	Intervention
Brand S., Gerber M., Beck J., Hatzinger M., Pühse U., Holsboer-Trachsler E.	2010	Intervention
Brassai, L., Piko, B.F., & Steger, M.F.	2011	Intervention
Brodáni J., Spišiak M., Paška L.	2015	Intervention
Brown D.R., Carroll D.D., Workman L.M., Carlson S.A., Brown D.W.	2014	Population
Burgess, G., Grogan, S. and Burwitz, L.,	2006	Population

Casey M.M., Harvey J.T., Telford A., Eime R.M., Mooney A., Payne W.R.	2014	Population
Castillo, I; Duda, JL; Alvarez, MS; Merce, J; Balaguer, I	2011	Population
Chatzisarantis, NLD; Hagger, MS	2007	Intervention
Chen L.H., Kee Y.H.	2008	Intervention
Chen, LH; Kee, YH; Chen, MY	2015	Outcome
D'anna C., Rio L., Paloma F.G.	2015	Intervention
Daniels E., Leaper C.	2006	Intervention
De Bruin A.P., Woertman L., Bakker F.C., Oudejans R.R.D.	2009	Intervention
Di Luzio S.S., Procentese F., Guillet-Descas E.	2014	Not available from the British Library
Doerksen, SE; Elavsky, S; Rebar, AL; Conroy, DE	2014	Intervention
Eime, R.M., Harvey, J.T., Brown, W.J. and Payne, W.R.	2010	Study Design
Findlay L.C., Bowker A.	2009	Population
Fløtnes I.S., Nilsen T.I.L., Augestad L.B.	2011	Intervention
Gardner,Sally May;Komesaroff,Paul;Fensham,Rachel	2008	Intervention
Gondoh, Y; Sensui, H; Kinomura, S; Fukuda, H; Fujimoto, T; Masud, M; Nagamatsu, T; Tamaki, H; Takekura, H	2009	Population
Hagensen K.P.	2015	Population
Hidalgo-Rasmussen C.A., Ramírez-López G., Martín A.H.-S.	2013	Intervention
Ivanović M., Milosavljević S., Ivanović U.	2015	Outcome
Jago R., Sebire S.J., Davies B., Wood L., Banfield K., Edwards M.J., Powell J.E., Montgomery A.A., Thompson J.L., Fox K.R.	2015	Population
Jalaludin B., Maxwell M., Saddik B., Lobb E., Byun R., Gutierrez R., Paszek J.	2012	Population
Jančiauskas, Rolandas	2012	Population
Jelalian, E; Hart, CN; Mehlenbeck, RS; Lloyd-Richardson, EE; Kaplan, JD; Flynn-O'Brien, KT; Wing, RR	2008	Outcome
Jonsdottir I.H., Börjesson M., Ahlborg Jr. G.	2011	Population
Kaczmarek,Lukasz D.;Drażkowski,Dariusz	2014	Intervention
Kallings L.V., Leijon M., Hellénus M.-L., Ståhle A.	2008	Population

Kantor,Rachel M.;Grimes,G. R.;Limbers,Christine A.	2015	Population
Karadağ Çaman Ö., Özcebe H.	2011	Intervention
Kardefelt-Winther, D	2014	Intervention
Kavetsos, G. and Szymanski, S.	2010	Study Design
Kelly,Nichole R.;Mazzeo,Suzanne E.;Evans,RonaldK.;Stern, Marilyn; Thacker,Leroy F. ;Thornton,Laura M.;Laver,Joseph H.	2011	Population
Kern M.L., Waters L.E., Adler A., White M.A.	2015	Intervention
Khan, Y; Taghdisi, MH; Nourijelyani, K	2015	Intervention
Kim J., Suh W., Kim S., Gopalan H	2012	Intervention
Kipp, LE; Weiss, MR	2013	Population
Knab A.M., Nieman D.C., Sha W., Broman-Fulks J.J., Canu W.H.	2012	Population
Knifsend,Casey Anne	2015	Population
Komlosi, E	2014	Intervention
Kort-Butler L.A., Hagewen K.J.	2011	Intervention
Kowert, R; Vogelgesang, J; Festl, R; Quandt, T	2015	Intervention
Lafrenière M.A., Vallerand R.J., Donahue E.G., Lavigne G.L.	2009	Intervention
Laure, P; Binsincer, C	2009	Population
Laure, P; Binsincer, C	2009	Population
Laurendeau J.	2014	Intervention
Lazaridou, A; Kalogianni, C	2013	Outcome
Le Menestrel S., Perkins D.F.	2007	Intervention
Lee A.J.Y., Lin W.H.	2007	Outcome
Lee B.W., Leeson P.R.C.	2015	Intervention
Leggett,Diane K.	2010	Intervention
Lestan K.A., Eržen I., Golobič M.	2014	Population
Lerversen, I; Danielsen, AG; Birkeland, MS; Samdal, O	2012	Study Design
Lieber S.B., Redberg R.F., Blumenthal R.S., Gandhi A., Robb K.J., Mora S.	2012	Population
Liu M., Wu L., Ming Q.	2015	Study Design – Systematic Review

Lopez-Walle, J; Balaguer, I; Castillo, I; Tristan, J	2012	Population
Lorger, M; Mrakovic, S; Hraski, M	2012	Population
Lu, FJH; Hsu, YW	2013	Intervention
Lupu E., Petrescu A.L.	2012	Study Design
Mack, D.E., Wilson, P.M., Gunnell, K.E., Gilchrist, J.D., Kowalski, K.C. and Crocker, P.R	2012	Study Design
Madison G., Paulin J., Aasa U.	2013	Population
Maffulli N., Longo U.G., Spiezia F., Denaro V.	2010	Intervention
Magnusson M., Hallmyr Lewis M., Smaga-Blom M., Lissner L., Pickering C.	2014	Study Design
Mäkinen M., Lindberg N., Komulainen E., Puukko-Viertomies L.-R., Aalberg V., Marttunen M.	2015	Population
Mancini, JA; Bowen, GL; O'Neal, CW; Arnold, AL	2015	Intervention
Martin-Albo, J; Nunez, JL; Dominguez, E; Leon, J; Tomas, JM	2012	Population
Maugendre M., Spitz E.	2011	Study Design
McDade-Montez, Elizabeth; Wallander, Jan; Elliott, Marc;Grunbaum,Jo Anne; Tortolero,Susan; Cuccaro,Paula; Schuster,Mark A.	2015	Intervention
McGee, R., Williams, S., Howden-Chapman, P., Martin, J. and Kawachi, I	2006	Study Design
McMahon E.M., Corcoran P., O'Regan G., Keeley H., Cannon M., Carli V., Wasserman C., Hadlaczky G., Sarchiapone M., Apter A., Balazs J., Balint M., Bobes J., Brunner R., Cozman D., Haring C., Iosue M., Kaess M., Kahn J.-P., Nemes B., Podlogar T., Poštuvan V., Sáiz P., Sisask M., Tubiana A., Värnik P., Hoven C.W., Wasserman D.	2016	Population
Medeiros M.D., De Castro Filho J.A.	2014	Not available from the British Library
Merrill R.M., Aldana S.G., Bowden D.E.	2010	Population
Mihaela, Cristuță Alina	2012	Population
Mochon, D; Norton, MI; Ariely, D	2008	Population
Mohan S., Smith C.A., Corriveau N.L., Kline-Rogers E., Jackson E.A., Eagle K.A., Goldberg C., Durussel-Weston J.	2012	Intervention
Molina J.J.M., Castillo A.S., De La Serrana H.L.G., Díaz M.Z.	2009	Population
Molina-Garcia J, Castillo I, Queralt A	2011	Study Design
Moljord I., Moksnes U.K., Eriksen L., Espnes G.A.	2011	Study Design

Moutão J., Alves S.M., Monteiro D., Cid L.	2015	Population
Nicholls L., Lewis A.J., Petersen S., Swinburn B., Moodie M., Millar L.	2014	Intervention
Noack P., Kauper T., Benbow A.E.F., Eckstein K.	2013	Study Design
Orkibi H., Ronen T., Assoulin N.	2014	Population
Papaioannou A.G., Appleton P.R., Torregrosa M., Jowett G.E., Bosselut G., Gonzalez L., Haug E., Ertesvaag V., Zourbanos N.	2013	Population
Peng W., Crouse J.	2013	Outcome
Pérez Ugidos,Guillermo;Laíño,Fernando,A.;Zelarayán,Julio;Márquez,Sara	2014	Intervention
Phillips G Renton A Moore DG Bottomley C Schmidt E Lais S Yu G Wall M Tobin P Frostick C Clow A Lock K Petticrew M Hayes R	2012	Population
Piqueras J.A., Kuhne W., Vera-Villarroel P., Van Straten A., Cuijpers P.	2011	Study Design
Proctor C., Tsukayama E., Wood A.M., Maltby J., Eades J.F., Linley P.A.	2011	Intervention
Pyky, R; Jauho, AM; Ahola, R; Ikaheimo, TM; Koivumaa-Honkanen, H; Manysaari, M; Jamsa, T; Korpelainen, R	2015	Intervention
Reding, Frank N; Grieve, Frederick; Derryberry, W. Pitt; Paquin, Anthony R	2011	Outcome
Riley A., Anderson-Butcher D.	2012	Population
Rössler R., Donath L., Verhagen E., Junge A., Schweizer T., Faude O.	2014	Study Design – Systematic Review
Rotheram-Borus M.J., Swendeman D., Becker K.D.	2014	Population
Sagatun, A., Sjøgaard, A.J., Bjertness, E., Selmer, R. and Heyerdahl, S	2007	Study Design
Sage, L; Kavussanu, M	2010	Population
Salehi, A; Harris, N; Sebar, B; Coyne, E	2015	Population
Schlarb A.A., Schwedler V., Feichtinger P.	2012	Study Design
Schmiedeberg C., Schröder J.	2016	Population
Schuch F.B., Pinto S.S., Bagatini N.C., Zaffari P., Alberton C.L., Cadore E.L., Silva R.F., Krueger L.F.M.	2014	Population
Schulz, KH; Meyer, A; Langguth, N	2012	Population
Schwanen, T; Wang, DG	2014	Population
Sekot A.	2013	Population

Sellakumar G.K.	2015	Intervention
Shaffer-Hudkins, Emily	2012	Population
Shiue, I	2016	Population
Sidoti E., Paolini G., Tringali G.	2010	Population
Sigvartsen J., Gabrielsen L.E., Abildsnes E., Stea T.H., Omfjord C.S., Rohde G.	2016	Study Design
Sjögren K., Hansson E.E., Stjernberg L.	2011	Population
Snyder A.R., Martinez J.C., Bay R.C., Parsons J.T., Sauers E.L., McLeod T.C.V.	2010	Study Design
Spandler H Mckeown M Roy A Hurley M	2013	Population
Spengler, Sarah; Woll, Alexander	2013	Population
Stein C., Fisher L., Berkey C., Colditz G.	2007	Population
Stenseng, Frode; Forest, Jacques; Curran, Thomas	2015	Population
Stubbe J.H., de Moor M.H.M., Boomsma D.I., de Geus E.J.C.	2007	Population
Sztankovics A.	2013	Population
Tanimaru J.H., Dos Santos A.L.P.	2016	Study Design
Tharenos C.L., Santorino D.	2009	Not available from the British Library
Thøgersen-Ntoumani C., Ntoumanis N.	2006	Population
Thomley B.S., Ray S.H., Cha S.S., Bauer B.A.	2011	Population
Vilela C., Gomes A.R.	2015	Intervention
Wall M., Hayes R., Moore D., Petticrew M., Clow A., Schmidt E., Draper A., Lock K., Lynch R., Renton A.	2009	Study design
Wicker, P; Coates, D; Breuer, C	2015	Population
Wicker, P; Frick, B	2015	Population
Williams K., Davis III O., Gittelman M., Pomerantz W.J.	2006	Population
Woodall, J; White, J; South, J	2013	Population
Yamada K., Kawata Y., Nakajima N., Hirose M.	2012	Outcome
Zook K.R., Saksvig B.I., Wu T.T., Young D.R.	2014	Outcome
Zullig, Keith J.; White, Rebecca J.	2011	Population

Appendix 3: Data Extraction Form

Title, Author, year		
Study objectives		
Study design		
Method of allocation to study group		
Measures of wellbeing (Include scale(s) used and time-points)		
Details of analysis (Include type of analysis i.e. quantitative/qualitative/mixed, and method and/or process of analysis e.g. thematic analysis/statistical analysis, any subgroup analysis and any methods used in the treatment of missing data)		
Participants included (at baseline and follow up in each group) (Source/recruitment, eligible and selected, number, age restrictions, gender)	Intervention	Comparator
Intervention(s) and comparison group(s) (Type, content, intervener, duration, method, mode or timing of delivery)		
Results (Key numerical results including proportions experiencing relevant outcomes in each group, means, medians, standard deviations, ranges and effect sizes with precision estimates e.g. confidence intervals/ p values whether or not significant [if P values are not reported this should be stated]. For qualitative data what categories/themes were found, results drawn by authors and evidence provided. Identify any inadequately reported missing data)		
Protected characteristics (methods and findings that relate to protected characteristics [age, sex, gender reassignment, sexual orientation, disability, race, religion, pregnancy/maternity, marriage/civil partnerships] and income and/or socio-economic status.		
Limitations identified		
Review conclusions (for each comparison made)		
Conflicts of interest and sources of funding		
Ethical procedures reported		

Grade/CERQual Rating	
----------------------	--

GRADE and CERQual for judging certainty / quality of evidence

Quantitative: Grade

Type of evidence	Randomized trial = high Observational study = low Any other evidence = very low
Decrease grade if (Each quality criteria can reduce the quality by one or, if very serious, by two levels.)	<ul style="list-style-type: none"> • Serious or very serious limitation to study quality (e.g. Important inconsistency; major uncertainty about directness; imprecise or sparse data; high probability of reporting bias)
Increase grade if	<ul style="list-style-type: none"> • Strong evidence of association—significant relative risk of > 2 (< 0.5) based on consistent evidence from two or more observational studies, with no plausible confounders (+1) • Very strong evidence of association—significant relative risk of > 5 (< 0.2) based on direct evidence with no major threats to validity (+2) • Evidence of a dose response gradient (+1) • All plausible confounders would have reduced the effect (+1)
Grade Rating / Range	High quality evidence Moderate quality evidence Low quality evidence Very low quality evidence

Qualitative: CERQual

Increase confidence if	<ul style="list-style-type: none"> • Study is well designed with few limitations • Evidence applicable to context (perspective or population, phenomenon of interest, setting) specified in objectives • Findings/conclusions supported by evidence and provide convincing explanation for patterns found • Data supporting findings is rich and good quality
Decrease confidence if (Each quality criteria can reduce the quality by one or, if very serious, by two levels)	<ul style="list-style-type: none"> • Serious or very serious limitations in design or conduct of the study • Evidence is not relevant to the study objectives • Findings/conclusions are not supported by the evidence • Data is poor quality and inadequate to support findings
CERQual Confidence Rating / Range	<p>High confidence It is highly likely that the review finding is a reasonable representation of the phenomenon of interest</p> <p>Moderate confidence It is likely that the review finding is a reasonable representation of the phenomenon of interest</p> <p>Low confidence It is possible that the review finding is a reasonable representation of the phenomenon of interest</p> <p>Very low confidence It is not clear whether the review finding is a reasonable representation of the phenomenon of interest.</p>