Impact of arts education on children’s learning and wider outcomes

ABSTRACT
Almost every educational system in the world regards numeracy and literacy as more important than the arts. In the UK arts interest groups and politicians have, for years, asked for arts education to be accorded the same priority arguing that arts contribute to learning and development of useful skills. It is not clear if these claims were based on evidence. The aim of this review is to examine international empirical evidence to establish whether arts participation could lead to improvements in children’s learning and other wider outcomes, and to identify arts activities with the best evidence of success. A search of 11 databases identified 76,195 reports, of which 200 were included for analysis. No studies with high levels of supporting evidence were found because of serious weaknesses in design. Tentative evidence suggests that music training and integrating drama in the classroom may have beneficial effects. This finding has to be taken with caution because of weaknesses in these studies, the lack of replication and inconsistent findings across studies. More robust and rigorous evaluations are needed to test the causal links. However, if improving attainment is the aim then arts may not be the solution. Promising programmes already exist that can boost learning. Given the lack of evidence so far, perhaps we should think more broadly about the purpose of arts.

Keywords: arts education, review, cognitive outcomes, noncognitive outcomes

INTRODUCTION
Arts have always struggled to find a legitimate place in the school curriculum in education systems across the world. For years, arts education has been marginalized and viewed by many as the poor relation of subjects such as maths, science and literacy, which are often regarded as core subjects because they are deemed to have greater currency in a competitive global economy (CBI, 2010, p.2).

International league tables (e.g. PISA and TIMSS) also used performance in these core subjects as an indicator of the quality of a country’s education system. Consequently, schools around the world are putting greater emphasis on these subjects. Performance in these curriculum subjects has now been used as a yardstick for attainment. To ensure that every child has the opportunity to succeed in education the No Child Left Behind Act was introduced in the US. This has led to a decline in emphasis on arts education. In the UK, reforms in education have also been made to ensure that the country remains competitive. These recent reforms and the focus on EBacc have been blamed for a decline in the number of students taking arts at GCSE (Telegraph 2014).

The marginalization of arts, the declining take up and the low status have prompted some to call for greater attention to be given to arts education in schools. Three main arguments were put forward. The first was that arts are synonymous with creativity, and creativity was necessary for a world-class education. For example, in 1999 the National Advisory Committee on Creative and Cultural Education (NACCE) argued that:
[...] no education system can be world class without valuing and integrating creativity in teaching and learning [...]

(NACCE 1999, p. 16)

The second argument was that arts encourage the development of important skills such as, critical thinking, motivation and self-confidence – skills necessary for innovation (House of Lords 2014).

Another argument was that arts facilitate learning at school, contributing directly to pupil attainment. In this sense they are relevant since they support pupils’ learning of the core subjects. In the UK there have already been movements to encourage arts participation in schools. For example, art and design and music are now taught in the primary school (DfE 2013). Arts specialist schools have also been introduced as part of a government-funded initiative. Pupils from disadvantaged backgrounds are offered music lessons in schools using the Pupil Premium Fund.

The aim of this paper was to review the empirical evidence worldwide to establish if arts participation leads to improvements in children’s cognitive and non-cognitive outcomes, and to identify the arts activities with the best evidence of success in supporting these outcomes. This review was commissioned by the Education Endowment Foundation to specifically look for causal evidence of the link between participation in arts activities and children’s learning and other wider outcomes with the aim of commissioning trials of the more promising approaches.

There is currently some indication that participation in creative activities may be beneficial (EEF Arts participation toolkit) to pupils’ learning and wider outcomes, but these are weak and inconsistent. A few promising studies have been published in the last decade in the area of music instruction (Schellenberg, 2004; Costa-Giomi, 2004) and of participation in the arts more generally (Smithrim & Upitis, 2005; Catterall & Peppler, 2007). Recent studies have looked in particular at the educational outcomes gained by at-risk pupils through music (e.g. Kraus et al., 2014) and arts enrichment programmes (e.g. Brown, Benedett & Armistead, 2010). Kraus (2014), for example, found that giving children regular group music lessons for five or more hours a week prevented an expected decline in reading skills for the poorer children. A longitudinal study reported that arts involvement was positively correlated with young people’s academic achievement (Catterall and Dumais 2012), but could not establish that participation was the cause of achievement. On the other hand, there are studies that suggest that arts programmes have little impact on academic performance at GCSE (Harland et al. 2000). A meta-analysis of studies on creative dramatics suggests that Arts education tends to be more effective for younger children than secondary school children (Conrad 1992). Another review on the impact of performing arts on secondary school pupils found very little research in this area that actually evaluated outcomes (Daykin et al. 2008).

These individual pieces of work, when examined, in isolation do not provide definitive evidence. It is the aim of the review to identify and synthesise all the available evidence to
get a more complete and coherent picture of the strength of impact of the different types of arts activities in schools.

This review was commissioned by the Education Endowment Foundation (UK) to specifically look for causal evidence of the link between participation in arts activities and children’s learning and other wider outcomes. The aim of this paper was to review the empirical evidence worldwide to establish if arts participation leads to improvements in children’s cognitive and non-cognitive outcomes, and to identify the arts activities with the best evidence of success in supporting these outcomes in order to make recommendations for future trials.

The arts activities that were of interest include the traditional fine arts such as visual arts, music, dance, performing arts, theatre and dance as well as modern dance and movement, hip hop, poetry and creative writing. Music activities can mean playing an instrument, singing, listening to music (classical or popular), or music as a curriculum subject. Visual arts include drawing, painting, sculpture, ceramic/pottery work and fabric printing. Arts education here refers to arts classes in school taken as part of the curriculum or integrated in lessons to support an academic subject. They may be taken outside school hours as part of the extra-curricular activity or as paid lessons (e.g. private tuitions). The outcomes of interest included both cognitive and non-cognitive or wider outcomes.

The review was conducted by a team at Durham University.

METHODS
The review began with a search of the eleven known educational, social sciences and psychological databases: (ASSIA, ERIC, BEI, International Bibliography of Social Sciences, ProQuest dissertations and theses UK & Ireland, ProQuest dissertations and theses Global, Social services and Sociological Abstracts, Educational Abstracts, PsycInfo and PsyARTICLES fulltext.

The search was limited to those reported or published in the English language, from 1995 to 2015. However, some older materials (pre 1990s) were picked up in the search. We kept some of these if we thought they might add to the evidence base. Only research reports on children aged 3 years to 16 in mainstream education were considered. Studies of interventions designed specifically for children with behavioural or learning disabilities and those who are in institutions or incarcerated were excluded. We considered only studies on arts programmes/initiatives or experiments that assessed impact on cognitive and non-cognitive outcomes. Opinion pieces, anecdotal accounts and promotional literature from interests groups were excluded.

The review was not intended to be comprehensive, but to identity arts activities that have the potential to improve the school achievement and wider/affective outcomes of young people from disadvantaged backgrounds, and to make recommendations for more robust trials in these areas. As is always the case with reviews of this nature which cover a wide range of subjects and outcomes, it is possible that some studies may be missed.
A broad search was conducted using the following keywords and their synonyms. These keywords were derived based on the brief given by the Education Endowment Foundation:

- art/music/drama/dance/fine art/creative writing/poetry or poem/
  verse/rhyme/limerick/creative (expressive/imaginative) writing;
- programme/initiative/project/education/training;
- trial/evaluation/intervention/trial/experiment/review/meta-analysis/causal/regression discontinuity/longitudinal/instrumental variables/randomized controlled trial/cohort;
- attainment/achievement/learning (school/cognitive/academic) outcome/critical thinking/key stage/exam/qualification/school readiness/non cognitive outcome (attitude/expectation/aspiration/behaviour/motivation/self-efficacy/ locus of control)

It has to be mentioned that the keywords here are not exhaustive, and cannot be exhaustive.

The search identified a total of 76,195 study reports (63,195 from the Social Science databases and 13,000 from the Psychological databases). These were imported onto EndNote. The entries were then screened for duplicates and relevance based on the title and abstracts (Moher et al (2009) by four researchers. A large proportion of these were duplicates. From the abstracts 244 were judged to be directly relevant to the topic and not duplicates of each other. In other words, they must be studies about any art activities (such as visual art, which includes painting, drawing, sculpture, pottery and fabric making), all forms of music from classical to modern pop), creative writing, poetry, dance (including classical, folk, modern, hip hop and street dance) and drama (this could be role play and performing art). Media and film studies were not part of the remit in this review.

To check that relevant studies were not inadvertently omitted (Torgerson 2003) one researcher then randomly screened 10% of the 76,195 studies After reading the full paper, forty-five were removed because they were not of the right age group, not empirical, not research or not for children in mainstream education. This left 199 which were included for data extraction. One paper was later included from the handsearches. This brought the total number to 200. Only empirical studies were considered. These included experiments, quasi-experiments, as well as observational pieces. The rest were discarded if they did not meet the inclusion criteria. Ethnographic studies, which are narrative accounts of the researchers’ experience as well as those that espouse theoretical frameworks of how arts education might be beneficial to learning, were also excluded. However, a few ethnographic narrative studies on creative writing and poetry were included because of the sparsity of studies in these two areas. The screening was the most time-consuming and labour-intensive part of the process.

The included studies were then assessed for their quality to judge the security or trustworthiness of the evidence (Gorard 2014). This was judged as high, medium or low evidence based on the research design (sample size, selection and allocation of sample, attrition, threats to validity and reliability and appropriate analyses). Two researchers
independently assessed and rated a sample of ten studies about music and ten on arts in general. One rated the music studies first and the other general arts. They then swapped over and compared their rating. There was very high agreement between the two researchers using the same reliability index (Gorard 2014).

A high quality study would be one which provides good evidence. A study is judged as high quality if it has a large sample (over 100 in each intervention arm), has low attrition (under 10%), involves random allocation of cases (or establishes baseline equivalence or uses pre-post comparisons), has low threats to validity (e.g. uses standardized tests, blinding, minimizes contamination or diffusion, includes a counterfactual) and reliability (i.e. large representative sample that can be generalized to wider population). The first three criteria are given the most weight. Studies that violate these three would be immediately regarded as weak. A small study which meets all the criteria would be regarded as medium (because of threat to reliability). In this review, no strong evidence studies were found. The best studies were judged medium or medium to weak.

We largely ignored the reported outcome in our assessment because there were several studies that reported positive impact even though there was no data. There were a number, which completely ignored the data and made their own conclusions independent of the data. These would be judged weak.

Studies that reported positive impacts and rated medium in evidence would have greater evidence of impact compared to those which reported positive impacts but rated weak in evidence. For example, there were reports that claimed that children had made significant gains at the end of the intervention without any comparison group. With no true counterfactual it is impossible to say if the children would have made similar progress if they had not been on the programme. There were studies that compared children in one cohort with those from previous cohorts in one school. Any differences could be due to differences in pupil intakes. There were also studies that compared children who were selected by teachers for participation with those that were not, and some compared children in arts-focused schools with non-arts focused schools. These studies would be judged weak in evidence as the two groups of children are likely to be different so any improvements could be due to differences in the type of pupils and not necessarily the result of the intervention.

It is also not uncommon to find studies that reported big gains on teacher/researcher-developed tests or teacher surveys. A number of studies also reported significant improvements based on teachers’ perceptions of impact, but no effects based on standardized tests. There were a number of large-scale correlational studies using national data. These were generally well conducted, and were rated medium, mainly because they suggest a link between arts participation and pupil outcomes, but on their own they cannot demonstrate causality. The strength of evidence reported, therefore, takes into account all these design and reporting issues.

The identified studies were then classified by:

- phase of schooling (pre-school, primary and secondary)
types of interventions broken down by subjects (e.g. music, art, dance and drama, creative writing and multi-arts).

Pre-school children refers to children under six years of age. Primary school refers to children aged six to 11 and the secondary school phase is for children aged 12 to 16.

In this paper we discussed separately those arts activities that show high levels of supporting evidence and those that have little evidence of impact. The latter refers to those that have no or negative effects or where effects were not evaluated, as well as those where there were too few studies to judge the strength of the evidence and the ones where the results are mixed or inconclusive.

Very few research reports were found on poetry and creative writing as a creative activity, and these were all poor. For the purpose of this paper we will, therefore, focus our discussion on the main arts activities.

RESULTS

The vast majority of the studies in the review were about music education and its effects on young people’s cognitive, social, emotional, behavioural and academic outcomes (Table 1). There were also proportionately more arts education studies for primary school children than any other age group (Table 1).

| Table 1: Arts activities and phase of schooling |

STUDIES THAT SHOW POTENTIAL

No studies which were judged as providing strong evidence was found. 72% of research reports on music indicated positive effects on a range of outcomes (Table 2). Music is the only art form that has the highest number of positive studies showing medium strength of evidence. This is not surprising given that 36% of all the studies were about music. However, not all types of musical experience are effective.

| Table 2: Music (n=72) |

Music training

Music training (for example learning to play an instrument) is the only art form that provides the best evidence of positive influence on the cognitive and non-cognitive outcomes of children across all ages.
For pre-school children positive effects were reported for creativity (Duncan 2007), spatial-temporal ability (Gromko and Poorman 1998), IQ scores (Kaviani et al. 2014; Nering 2002), reading and language (Myant et al. 2008; Harris 2011). An experiment conducted on 10 sets of monozygotic twins (Nering 2002) aged 3 to 7 showed that the twin who received private piano instruction improved in IQ and arithmetic scores while the other twin who received no training did not show improvement. This is a very small study and so the findings must be treated with caution.

For primary school children, positive effects were reported for speech (e.g. François et al. 2013; Moreno et al. 2009), brain development (e.g. Olson 2010; Schlaug et al 2005; Schellenberg 2004; Degé et al.), academic outcomes (e.g. Harris 2008; Piro and Ortiz 2009) and other cognitive skills (e.g. Roden et al. 2014; Costa-Giomi 1999).

Some studies also suggest that playing an instrument can improve secondary school aged children’s self-concept, self-efficacy, motivation and behaviour (Degé et al. 2014; Wilson 2009; Rodgers 1999; Fitzpatrick 2006; Kurt 2011).

However, the evidence is not strong enough to provide conclusive results. Only eight of the positive studies were rated medium/medium-low in quality. The rest were rated weak for various reasons. One study (Harris 2008), for example, compared the maths achievement of children receiving the conventional Montessori education with those using a music-enhanced instruction. This involved only 190 children from one type of school thus limiting the generalization of the findings to other populations. Crucially, there was no comparison of gain scores, and it was not clear if children were randomly allocated. The higher scores achieved by the music group may suggest that the two groups were not equal to begin with.

A recent trial in the UK (Act, Sing and Play), which replicated Schellenberg’s study, found that disadvantaged children receiving music instruction (playing string instrument and singing) did not perform better in maths and literacy compared to children who were taught drama (Haywood et al. 2015). This study involved 909 Year 2 pupils (age 6/7). However, as the study did not have an inactive control group, it was not possible to say if these children would have done better without music and drama. Since music and drama are both arts activities, the study is unable to provide evidence of the impact of arts.

Other studies using brain scans also could not provide evidence of impact. Two studies showed that although there were changes in the brain development of musically trained children, there was no evidence that the changes translated to improvements in academic attainment (Schlaug et al. 2005; Olson 2010). Another study using brain scans of children of different ages with varying years of musical experience showed positive correlation between music training and visuo-spatial and motor co-ordination (Hudziak et al. 2014). The correlational design of the study could not rule out the influence of maturation and other confounding variables.

There were some well-conducted studies, but these tended to be small scale (e.g François et al. n= 28; Moreno et al. n=33), and the larger scale studies were invariably correlational and
thus could not establish causality (e.g. Moinar 2012; Degé et al. 2014; Wilson 2009; Rodgers 1999; Fitzpatrick 2006; Kurt 2011). Several compared students who opted for music as an exam subject with students who did not (Cabanac et al. 2013; Horton 2012; Van der Vossen 2013; Waller 2007). These studies can only suggest a link between music and outcomes, but cannot determine the direction of causation. Students enrolled in music at school may be inherently different to those who did not. For example, a large-scale analysis of standardized test scores in English showed that pupils who participated in school music programme performed better than non-music pupils (Johnson and Memmott 2006), but the effect was small and the influence of confounding variables could not be ruled out. Another large-scale analysis of standardized test scores of 15,630 high school students in the US found no differences between students who did music at school and those who did not (Elpus 2013) after controlling for background variables. Experimental studies were also weak as they were either small or did not randomly allocate participants to conditions (e.g. Strait et al, 2013). A meta-analysis of experimental studies provided tentative evidence of impact of music training on spatial-temporal reasoning, but only 5 of the 19 studies randomized participants to intervention conditions (Hetland 2000).

On the other hand, there were studies that indicated no or mixed effects of music. For example, one study involving 29 children (Charles 2014) found no differences between pupils who participated in music (instrumental music education in school) and those who received no music instruction in school on Graduate Exit Exam. Another study suggested that there was no relationship between school band involvement and social-emotional competence, but the analyses provided no evidence for this (Chase 2012). This study had only 37 pupils and 7 band members. Deer (2010) compared students in an outstanding music school with a non-music school. Positive effects were found for 4th grade children (age 9 to 10) on both reading and maths, but for 8th grade (age 13-14) pupils effects were found only for reading. A small-scale quasi experiment found mixed effects of keyboarding on students’ self-efficacy (Previti 2003).

Individually the evidence from these studies may be weak, but taken together the large number of positive studies suggest there is potential in this area that is worth pursuing.

**Integrating music in the curriculum**

There is some suggestive-indicative evidence that integrating music in the curriculum can facilitate learning, but only for primary school aged children. A number of studies have shown that integrating music in the curriculum can have positive effects on primary school children’s reading (e.g. Cochran 2009; Lyons 2009; Register et al. 2007; Peters 2011; Bryant 2013; Walton 2013) and maths (An 2013; Courey et al. 2012;). Two studies showed no effects, but both were rated weak. One had only 25 children taken from one school (Tendall 2010). The music and movement group already achieved the highest scores in the pre-test, leaving no room for improvements. The lack of difference between groups could be due to the ceiling effect. Another study reported that the groups were not statistically different on pre-post-test of reading attitude (Kingsriter 1998). The use of significant tests on such a small sample (n=2 classes) was also inappropriate.
However, there is little evidence that integrating music in the school curriculum is beneficial for the other age group of children.

**STUDIES WITH LITTLE SUPPORTING EVIDENCE OF EFFECT**

**Integration of music in the curriculum**

Several studies looking into the integration of music in classroom lessons reported positive impact on academic outcomes of pre-school children. These were methodologically weak. One involved only 14 children (McDonel et al. 2013) with no comparison group, using a convenience sample of children of highly educated parents. Another included only children whose parents were academics in a University child care centre (Ritblatt et al. 2013). Others did not involve random allocation of participants (e.g. Runfalo et al. 2012; Wellman 2007: Fisher 2011), or based results on classroom observations of a small number of children (Wade 2011). In Wade’s study positive changes were observed in only 3 out of the 8 children and the results were not convincing. Wellman (2007) reported higher post-test scores for experimental children, but it was not clear if experimental groups made bigger gains than control groups between pre and post-test. There were substantial differences in pre-test scores which could explain the higher post-test scores of the experimental group. The very small samples also had implications for external validity.

**Listening to music**

While playing an instrument has shown to have some evidence of impact on children’s learning and wider outcomes, listening to music or the so-called “Mozart Effect”, on the other hand, does not. Three studies were found; one suggested improvements in psychomotor skills (Thompson 2005) and one showed negative effects on memory. Those exposed to music performed worse than those without music (Bressler 2003). Both evaluations were flawed, involving very small sample (under 100) taken from one setting. This posed a threat to external validity. Bressler’s study was only a 10-minute session, too short for real effects (if any) to be realized. There was also high attrition from both groups. Thompson’s study did not involve random assignment of children. Children could be different in ability, motivation, prior musical experiences etc. The classes were also taught by different teachers – teaching quality and teacher characteristics may have influenced the results. A third study of 102 children aged 8 to 11 (Albright 2012) showed that listening to classical music had negative effects on children’s maths performance. Children who were randomly assigned to treatment condition listened to baroque and classical music for 50 minutes five days a week during maths lessons. These lessons were structured to include an element of music, for example, multiplying musical notes, using time signatures to learn fractions. The state standardised test taken 16 weeks later showed that control children did better in the maths test.

**Background music**

There is also no evidence that background music facilitates learning. In fact it may have detrimental effects. One study involving 334 students 7th and 8th grade (age 12-14) students
exposed students to 2 conditions: non-music environment and listening to music from Billboard Magazine’s (2006) top hit singles. Students’ reading comprehension performance declined significantly when listening to music. Those who expressed a strong preference for listening to music while studying also did particularly badly on the comprehension test. The findings of this single study alone cannot demonstrate that listening to music is detrimental. The evidence however is weak. There was no comparison group, and the kind of music played may be a factor rather than music itself. Different types of music may produce different results. It is also conceivable that children respond differently to music according to their learning styles. These were not tested. Nevertheless, there is not enough evidence to suggest that listening to background music in learning is a good thing.

A second study looked at the use of background music in three primary classrooms in England and its effects on children’s behaviour and attainment. The results were inconclusive. Generally, children performed better in reading with background music than without, but the reverse was true for maths performance. Effects on behaviour were unclear. Only 47 of the possible 90 pupils participated in the study. These were children whose parents volunteered. It was not clear what type of music was played in the study.

Visual arts

Table 3: Visual arts (n=26)

73% of studies on visual arts were rated weak. Four of the medium quality studies showed mixed results, and two medium studies actually showed no effects. Only three studies looked at the impact of visual arts for pre-school children. One did not evaluate outcomes (Hardy 2011), and the other two showed mixed results (Burger & Winner 2000; Borman et al. 2009). For example, positive effects were found for developmental reading and reading readiness, but not on phoneme skills and dictation and reading in general. This suggests that the results are sensitive to the kind of test and the elements tested.

Twelve (46%) of the studies on visual arts were for primary school aged children. There is also no evidence from these studies that visual arts can enhance the cognitive and non-cognitive outcomes of primary school children. Most of studies showed mixed results. The three positive studies all reported effects on cognitive outcomes (history, creative thinking and maths) and all involved integrating visual arts in the lessons. However, all had either a very small sample or unclear sample (Phillips and Bickley-Green 1998), allocated by class to treatment groups. In one study the experimental children also received additional enrichment activities which were not available to the control group (Brugar 2012). Two of the four positive studies did not have a control group (Fountain 2007; Phillips and Bickley-Green 1998). Finnan-Jones’ study reported positive effects on standardised tests of maths for 194 English language learners, but children were allocated non-randomly by class (n=4). Luffing’s (2000) study of SPECTRA+ (a creative arts programme) showed positive effects of SPECTRA+ on overall test of creativity but not on the subtest of elaboration. Positive effects
were seen in one district on maths, reading and vocabulary, but not in another district, and among boys and only for some grades. Different assessments were also used in different districts which could explain the different outcomes, and conditions were nested within schools so there is a possibility of diffusion. Catterall and Peppler (2007) found effects of visual arts on only some measures of non-cognitive outcomes. Another study found no effect of visual arts on standardized tests of non-verbal intelligence (Stephens 1996). These mixed results therefore present inconclusive evidence of impact.

Other studies showed that visual arts may even be harmful. One study involving 400 children (Huizenga & Van der Wolf 1996 in Haanstra 2000) indicated that arts as an extra-curricular activity in school had negative effects on primary school children’s reading and arithmetic (measured on standardized tests). Standardised tests on non-cognitive outcomes (achievement motivation, self-image and peer acceptance) also showed no obvious effects of participation.

An evaluation of a school-based creative arts programme (Wurzel 2012) reported negative effects on problem-solving behaviour. Although experimental children made bigger gains than control children on self-esteem and affective outcomes, the differences were not significant. Because of the very small sample (n=29) and lack of randomization the evidence of impact is therefore not reliable. Groups were not equal to begin with. 85% of control children had anger issues compared to 38% of experimental children. The intervention was also not fully implemented because of non-compliance within groups due to disruptive behaviours.

Another study suggests that teaching children aesthetic appreciation of art does not improve children’s metacognition and literary skills (Dennis 1995). This was a small study (n=52) and children were not randomly selected nor allocated to groups. Both treatment groups were exposed to aesthetic appreciation prior to the study, so there was diffusion of intervention effects. There were no reliable measurements for assessing literary analysis. The assessments were specific to the instruction. The intervention involved getting pupils to practice making a concept map addressing the question: What do you look for and think about in a work of art? The test prompts used exactly the same practice questions that were familiar to the experimental pupils but not for the control children.

The evidence for visual arts on secondary school aged children was also inconclusive. A Turkish study suggested beneficial effects, but provided no information about the tests used nor about the results (Kalyonocu and Tepecik 2010). Another study using arts-related ICT showed no effects on children’s problem-solving skills (Liem et al. 2014). A large-scale longitudinal study involving 2,906 pupils was unable to show if it was participation in fine arts that contributed to higher grades or that pupils who took up fine arts were different from those who did not. It is also possible that they were more likely to be perceived as intelligent, and hence received more attention and assistance from teachers (Di Maggio 1982). Two reported mixed results (Alo 2009; Schultz 2011), and a further three showed no beneficial effects (Ben-Chetrit 2014; Webb 1985; Nagel et al. 1997). In fact Nagel et al. suggested that art education in secondary school may have a negative impact on exam
grades and later life outcomes compared to those doing music or not doing arts. These findings are not tenable because of the high percentage of missing data. The correlation study also shows only association between interest in art and participation in fine arts, but not causation. Caution has to be exercised when interpreting such results.

Dance

Table 4: Dance (n = 15)

The evidence for dance is mixed. Only two medium quality studies were found; one showed positive effect on reading skills for primary school children (Rose 1999), while another showed no effect (Zander et al. 2014) on peer relationships of children aged 10-15. The others were all rated weak. A meta-analysis showed a small overall effect of dance on reading and non-verbal reasoning, but the results were inconclusive because of the small, heterogeneous samples (Keinanen, Hetland and Winner 2000).

Only two studies were found for pre-school children. One did not evaluate outcomes (Cheung 2010), and the other was based on teacher/parents’ perceptions (e.g. Lobo and Winsler 2006).

Six studies on dance were found for primary school children; two of these reported negative impact. Disanto-Rose (1996) found that children exposed to creative dance did worse than control children on spatial concepts and spatial awareness. However, there was a possible dilution of effects as the control classes were exposed to square dancing. Von Rossberg-Gempton (1998) also reported negative effects of creative dance on children’s psychomotor skills and self-concept. This was a very small study with fewer than 20 in each group and groups were not randomly allocated. Both studies were therefore rated weak. One study did not evaluate the impact but assessed children’s cognitive process in creating dance (Giguere 2007), and one reported effects on only some aspects of academic self-concept for girls only (Stratton-Gonzales 2008). Riley (1984) suggested positive relationships between creative dance and body image, self-esteem and problem solving skills. However, performance on problem-solving skills was not tested, and improvements in body image were based on pupils’ self-reports. Rose (1999) reported positive impact of Whirlwind’s basic reading programme on primary school children’s reading skills. This was a well-designed study with a large sample (n =362) from 12 schools. However, schools were not randomly assigned to treatment and the high pre-test scores of the dance groups suggest that the groups were not balanced at the outset, so the bigger gain scores could be a case of regression to the means.

Of the five studies for secondary school children, only two reported positive effects; one on creative thinking (Minton 2000) and one on stress (Roberts 2010). Roberts’ study used raised body temperature as an indicator of reduced stress after dance activity. Applying the instrument immediately after physical activity like dancing may not be valid as such activity is likely to result in raised body temperature. There was no comparison group, so it was not
possible to compare the effects with similar other activities. This was a very small study involving only 10 self-selected individuals. Bridley’s (2014) study showed no effects of aerobics dance on girls’ self-efficacy and their level of physical activity. Participants were taken from one school, and there was no comparison of gains between pre- and post-dance programme. Another study suggested positive effects of creative dance on creativity but not critical thinking (Kim 1998). Only two classes were involved and the researcher taught both experimental and control class. There were serious issues with the validity of the test. Possible contamination was introduced as half the class used Form A and the other half Form B for the pre-test. This was then switched over at post-test. The test of critical thinking had no parallel version, so the same test was used for both pre- and post-test, which was taken only two months apart. There was therefore a practice effect.

**Drama**

**Table 5: Drama (n = 27)**

14 of the studies on drama (53%) suggested positive effects, but only one of these was considered medium in quality. The other medium quality studies suggested either mixed or no effects of drama on outcomes (Table 5).

Three studies for pre-school children suggested beneficial effects on reading/decoding text and story recall (Goodman 1990; Pellegrini and Galda 1982; Pellegrini 1984), but the evidence was not strong. They were all small studies with fewer than 50 in each intervention group. One did not have a control group and two did not have pre-post-test comparisons, so it was not possible to say if improvement in reading skills was due to the drama activity or natural development of the children.

For primary school aged children creative drama also shows promise on the academic outcomes though not for non-cognitive outcomes. Five reported positive effects, and four showed effects for some measures but no effects on others (Goldstein 2010; Laurin 2010; Rose 2000; Fizzano). In most cases creative drama was integrated in the classroom as part of an instructional strategy. Joseph (2014) suggested that creative dramatics could help with the vocabulary development of young children. However, the researcher-teacher developed instrument could bias assessment as the researcher might inadvertently pick vocabulary words that were taught in the creative dramatic lessons. Schaffner et al. (1984) suggested and speculated on the possible effect, but did not evaluate the impact. As there was no comparison group it was not possible to say if drama activities had any beneficial effect on language use compared to regular classroom activities. Roberts (2007) reported no effect on self-concept, but there were several issues with the study design. Children were non-randomly selected and allocated and both groups were exposed to other forms of art, music and other innovative teaching instruction as well. All these could have diluted the effects. Freeman et al. (2003) also found no effects of creative drama on self-concept and other non-cognitive skills (social skills and behaviour). This was a well-conducted study with proper randomization and pre-post-test comparisons. Although the sample was reasonably large (n=237), it was divided into four groups. Effectively, there were only 49 participants in the
treatment group and 47 in the control with pre- and post-test data. Since all the children received some kind of enrichment activities (music or drama) the benefits of the treatment may be obscured if participation in music influences the traits related to the outcomes measured. Haywood et al. (2015) also compared children who received training in music (playing string instruments), singing and drama. No differences between groups were found. This could be due to diffusion of effects.

One study suggested that creative drama as an enrichment activity might have an adverse effect on children’s behaviour (Danner 2003). The evidence, however, was not reliable given that there was only one study with a small sample (n=54). The groups were not equal as the treatment pupils were volunteers. The question items used in the assessment were similar to those addressed in the intervention. It is possible that the intervention encouraged pupils to be open and frank about their intention to take drugs or alcohol, hence the negative results.

Six out of the eight studies for secondary school children suggested that creative drama could have beneficial effects on academic outcomes (Catterall et al. 1999; Otten et al. 2004; Duatepe-Paksu and Ubuz 2009; Arieli 2007; Cormack 2004). One showed positive effects on empathy and theory of mind (Goldstein 2010). However, because of methodological weaknesses in these studies, the evidence is rather weak. In some studies assignment to conditions was unclear (Cronack 2004; Duatepe-Paksu and Ubuz 2009). Goldstein’s study involved volunteers and no comparison of gain scores. It was therefore not possible to say if the control pupils would have made the same gains. The effect on non-cognitive outcomes was inconclusive. Daykin et al. (2008) reviewed three studies on performing arts and the conclusion was mixed. The first study conducted in UK measured self-concept using pupil ratings of multi-dimensional self-concept and teacher reports. The second study based in Canada used validated tests but the effects on social skills were inconclusive. A second Canadian study showed no effect on behaviour and peer relationships based on teacher reports.

Kratochvil (2006) reviewed studies on integrating drama in the curriculum, but could not confirm the causal effect on literacy due to design flaws of these studies and the limited number of empirical work. Lack of objective measurements of outcomes was another issue with the studies reviewed. A second study provided mixed results; positive effects on some measures of literacy but no effect on vocabulary development (Podiozny 2000). The quality of the studies reviewed was also questionable. There were no consistent measurements of outcomes resulting in widely different effect sizes reported. Weaknesses in individual studies were also not identified. It was therefore difficult to judge the quality of these studies.

Multi-arts activities

Table 6: Multi-arts (n = 38)
A large number of studies examining the effect of integrating a combination of arts activities in school reported positive results, with many rated medium in quality (see Table 6). The majority of these were large-scale correlational studies using national datasets. However, the correlational nature of these studies and the fact that the kind of arts was not specified makes it difficult to recommend such activities in school.

There were a couple of interesting programmes for pre-school children, but the evidence was weak. Two evaluations of the Kaleidoscope programme were reported in Brown et al. (2010), one did not have a true counterfactual and the other had a sample of only 63 children using a placebo (alternative programme) to control for the Hawthorne effect. What is impressive are the large effect sizes reported for academic outcomes (receptive vocabulary and early learning) measured using standardised tests (ES =1.7; 1.5)

Five studies on the effects of multi-arts were found for secondary school children. Four reported positive effects. Again the evidence was weak either because the studies did not include a counterfactual or did not randomize cases or both. One had no control group and no pre- post-test comparisons to establish changes in attitudes (Poe 2000). The researcher was also the teacher who also collected the data. This could have influenced participants’ responses. Another study (Konrad 1999) based outcomes on teachers’ and pupils’ self-report. Two other studies were by the same author (Catterall et al. 1999; Catterall 2012). Both were longitudinal studies comparing students’ level of arts participation and their academic performance. There was also no differentiation between the different art forms, so it was not clear which art activities had the most influence.

Although the longitudinal studies (e.g Catterall & Dumais 2012; Liem et al 2014) showed a strong association between arts participation and academic performance, these studies should not be taken to mean that arts participation lead to higher academic outcomes because in the US, high performing pupils were encouraged to study the arts.

There were also a number of meta-analyses investigating the effects of integrating a combination of arts activities on children across age groups. Most of these were conducted by the same authors (Hetland and Winner 2001; Winner and Cooper 2000; Vaughn and Winner 2000; Catterall 1998).

These meta-analyses produced mixed. In one review only two of the 18 studies with a strong or moderate causal design reported positive effects (Robinson 2013). Many of the studies rated as having strong causal design were small scale, had no random allocation and no control groups. There was also a possibility of conflict of interest as the author also wrote and developed programmes in arts integration.

Another meta-analysis of 31 studies suggested positive effects of multi-arts experience on maths and verbal skills, but the results failed to reach statistical significance on some tests (Winner and Cooper 2000). Results from experimental studies were inconclusive. One limitation of such meta-analyses was the lack of standardisation of outcomes across the range of studies. There was also little information about the individual studies. We had little
information about the kind of art forms and whether effects differed with different art forms and different age groups.

Although Vaughn and Winner (2000) concluded that arts participation and maths and verbal SAT [standardized test] scores were strongly correlated, effects on children for different age groups were not explored. It was not clear which art activities were beneficial for which age group of children. Winner and Hetland (2000) found no causal links between integrated arts activities and reading, maths and verbal reasoning, but medium causal link between music and spatial-reasoning.

Other large-scale correlational studies also reported positive associations between arts participation and cognitive outcomes (e.g. Catterall 1998; Burton et al. 2000). These studies compared the levels of students’ arts exposure with their academic outcomes and other wider outcomes. The positive associations, however, do not suggest causal relationships.

Five other reports showed mixed results for integrating multi-arts in school. One showed positive effects for young children but not for the older ones (Arthington 2002), another showed effects on self-regulation and reading only for at-risk children, but no effects on maths performance (Baum and Owen 1997). Both studies were flawed. Arthington did not involve a counterfactual so it was difficult to say if similar pupils would have made the same progress without the intervention. Reading grades were assessed by teachers but assessments were found to be inconsistent between teachers. Baum and Owen compared non-randomly selected children on an art talent programme with those who were not in any arts programme. It was also not clear what art forms were integrated in the lessons. In another study on after-school fine arts programme the author concluded that it was effective in raising the self-esteem of disadvantaged children, but had no effects on behaviour (Rossini 2005). One study was so badly described that it was not clear what the findings were (Martin et al. 2013).

On the other hand, other studies reported no beneficial effects of integrating fine arts on academic outcomes (Garcia 2000; Boyes & Reid 2005; DeMoss and Morris 2002; Thomas & Arnold 2011). Garcia (2000) and Thomas & Arnold (2001) reported no overall effects on integrating fine arts on reading and maths. Both were rated weak in evidence. There was no random allocation to groups and no baseline equivalence was established. Using post-test only design with no random allocation of subjects seriously undermine the validity of the results. In Garcia’s study comparisons were made with different cohorts of pupils. There was also a high mobility of children in and out of the school, as well as many incomplete or unreturned forms. Thomas and Arnold (2011) compared specialist arts schools with non-arts schools. Any differences could be due to pupil intake or school effects rather than the arts activities.

Almost all the studies in these meta-analyses were correlational, comparing arts-focused schools with non-arts schools. The findings from these studies therefore, could only be indicative since high performing schools were encouraged to provide for arts classes and so were able to retain their arts programmes which the lower performing schools were unable
to do. Arts-rich schools were also more likely to be progressive schools with innovative teachers and programmes. So the better academic performance could not be attributed to arts participation alone. In the UK where this was not the case, arts students actually performed worse in English and Maths in national exams (Harland et al, 2000) than non-arts students after controlling for prior attainment and social background. However, pupils reported improvements in creativity, critical thinking, self-confidence and social development. This suggests that self-reports may not be reliable and that improvements in well-being do not necessarily lead to better academic performance.

Comparing outcomes of high arts performing schools with low performing non-arts schools is not a fair comparison. Arts-focused schools also appealed to different types of pupils. It is also possible that integrating arts in the curriculum makes learning fun and enjoyable for pupils and teachers. Further research using an experimental design is needed to test these hypotheses.

**LIMITATIONS**

As with all systematic reviews of this nature which cover a wide range of subjects and outcomes, some studies may have been missed. For example, we did not search all the specialist arts databases. The parameters set, including only articles reported in English, from 1995 to 2005 means that some relevant studies outside these parameters were not included. Creative writing and poetry are topics normally considered as part of linguistics, literacy or language, and not arts as such. Since we did not include literacy or language in our key words, not many studies in these areas were found in our search. Also, as we did not specifically searched for terms like the “Mozart Effect”, there may well be more of such studies that were not found in our search. If time permits, a second review would be warranted. The key issue is whether including the studies that were missed would have altered the findings. To test this we did a quick search of the Web of Science on the Mozart Effect. This showed up 335 results. Thirteen were found to be relevant and met the inclusion criteria. The full text to a number of these studies could not be obtained. Our evaluation was therefore based on the abstracts, some of which did not specify the age of the participants but were included for analysis nevertheless. A large proportion of these were about the use of music for treating special conditions, such as epilepsy, Alzheimer’s and Parkinson’s disease or for participants outside the specified age (for example, pre-term babies and the elderly) or experiments performed on rats. The relevant ones were correlational studies, with no random assignment to treatment conditions (Sanders 2012; Perlovsky 2013). Perlovsky’s study showed that students who like or perform music have better grades in all subjects. This suggests that music students are inherently different to non-music students, but does not necessarily suggest that music causes students to perform better. One experiment (Hsieh 2011) comparing three groups of participants (Bach, Mozart and Silence) on the effects of music on spatial ability, found no differences between groups. This was a small study of only 90 participants. Since the full paper was not available we could not tell from the abstracts the age of the participants. Similarly Crnec et al. (2006) also found no effects of listening to Mozart on the spatial temporal performance of 136 ten to eleven year old children who were randomly assigned to three conditions (Mozart, popular music and silence). Schellenberg and Hallam’s (2005) randomised controlled trial also could not provide evidence for the beneficial effect of listening to Mozart. A matched comparison study (Lints and Gadbois 2003) showed that other conditions and not just listening to Mozart could explain enhancement in spatial reasoning.
On the other hand, there were studies reporting positive effects of listening to Mozart. One experimental study (Jausovec et al. 2006) suggested positive effects on spatial-temporal ability. This study reported two experiments, both involving very small numbers (one had 14 participants in each arm of the intervention and the second experiment had only 12). It was also unclear how the participants were allocated to the treatment conditions. Polzella (2010) also suggested positive effect of Mozart on cognitive performance. The full paper was not available, so it was not possible to assess the quality of the evidence. Another relevant study (Popp 2010) was a literature review on the Mozart effect in general. However, a meta-analysis (Voracek 2008) revealed publication bias, lab differences and nonspecificity in the Mozart effect. All these suggest that the evidence for the so-called Mozart effect is unsubstantiated. Therefore, including the Mozart Effect in our search would not have altered our findings.

CONCLUSIONS

Despite the huge body of work on arts education, little can be concluded about the beneficial effects of arts on children’s learning and wider outcome. This is largely due to weaknesses in the research, lack of replication and inconsistent findings across studies. However, as Sir Kevin Collins, the chief executive of the Education Endowment Foundation said, the absence of evidence does not mean the evidence of absence.

Although no studies with a high level of supporting evidence were found, there is one arts activity which potentially could facilitate children’s learning. Music training in different contexts has been studied repeatedly with many studies showing positive effects on both academic and non-academic outcomes for children of all age groups. However, the large-scale studies were largely correlational and conducted in the US using state-wide datasets. There were a few experimental studies, but these were either small-scale, lack randomization or had high attrition. This is an area where large-scale randomized controlled efficacy trials could be conducted to test the causal effects of music education.

A number of studies also suggest that integrating drama in the classroom could support learning but only for primary school children. No effects were found for non-cognitive outcomes. These studies could be replicated using rigorous and robust evaluation designs to confirm the causal effects.

There is currently very little research on arts education for the pre-school phase (aged 5 to 7). This suggests that there is a gap in research for this age group of children. For example, there were no experimental studies on the effects of poetry or rhymes on the development of vocabulary and reading fluency for pre-school children. There were also very few studies looking at the effects of creative writing as a tool for cognitive development and critical thinking of school aged children. These two areas are worth exploring with pilot trials before testing them on a larger scale.

How can research in this area be improved?

• First more rigorous and robust evaluations of the impact of arts activities are needed.
  These should ideally focus on one or two art forms so that effect of the specific art
activity can be isolated. They should be large scale (over 100 in each intervention arm) and involve random selection and allocation of samples to treatment groups to overcome the vagaries of confounding variables.

- There should be replications of the better positive studies to confirm the effects.
- Where pupils’ art products are assessed, judgements should be by independent expert judges who are blind to treatment allocation.
- As arts are often associated with fun and leisure activities, to make sure that any changes in outcomes are not due to the novelty effect, it is necessary to have an alternative innovative treatment for comparison (but not one that is similar to arts), in order to separate the effect of the arts activity from the fun aspect.
- Future research needs to think of more objective ways to measure non-cognitive outcomes.
- Rich measures should be developed to help evaluate whether arts can lead to the transfer of skills to other curriculum subjects.
- If interventions are to be carried out in the classroom by regular classroom teachers the teachers need to be trained accordingly, and willing to use the proposed strategies. Resistant from teachers can affect the successful implementation of the programme.
- Professional artists or certified drama or music teachers can be engaged who can incorporate, for example, mathematical concepts in their art lessons. These professionals can work together with classroom teachers to achieve the lesson objectives.
- Process evaluations are needed to understand the transference of arts learning or arts activities to other domains of learning (such as literacy or maths), and more importantly to understand the mechanism – and how this can be tested in future research.

The wider picture

Given the lack of evidence so far on the impact of arts on academic attainment, perhaps we should think more broadly about the purpose of arts. If improving children’s attainment is the aim, then perhaps we should look to other approaches since promising approaches already exist for teaching literacy and numeracy. Arts should be enjoyed for its own sake. If children enjoy arts and if that leads to improvements in academic attainment and other affective outcomes, that is a bonus, but it should not be the main purpose for having it in the school curriculum.

ACKNOWLEDGEMENTS

The review on which this paper is based was funded the Education Endowment Foundation.

REFERENCES


**STUDIES USED IN THE REVIEW**


Harris, D.J. (2011) Shake, rattle and roll - can music be used by parents and practitioners to support communication, language and literacy within a pre-school setting? *Education 3-13: International Journal of Primary, Elementary and Early Years Education, 39*(2): 139-151.


Pacheco, D.L. (2011) *Writing out loud: The intellectual and epistemological work of*


Schaffner, M., Little, G. and Felton, H. (1984) *Nadie Papers No.1, Drama, language and learning: reports of the drama and language research project, speech, drama*
centre. Tasmania: Education Department of Tasmania. National Association for Drama in Education. Education Department of Tasmania.


### Table 1: Arts activities and phase of schooling

<table>
<thead>
<tr>
<th></th>
<th>Pre-school (under 6)</th>
<th>Primary (6-11)</th>
<th>Secondary (12-16)</th>
<th>Across age groups</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual art/general art (drawing, painting, sculpture, pottery)</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>Drama</td>
<td>5</td>
<td>12</td>
<td>8</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>Dance</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Music</td>
<td>17</td>
<td>31</td>
<td>20</td>
<td>4</td>
<td>72</td>
</tr>
<tr>
<td>Creative writing</td>
<td>-</td>
<td>1</td>
<td>6</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>Poetry</td>
<td>-</td>
<td>5</td>
<td>10</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>Multi-arts</td>
<td>1</td>
<td>13</td>
<td>5</td>
<td>19</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>80</td>
<td>62</td>
<td>30</td>
<td>200</td>
</tr>
</tbody>
</table>

### Table 2: Music (n=72)

<table>
<thead>
<tr>
<th></th>
<th>Not determined</th>
<th>Weak</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>No beneficial impact</td>
<td>1</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Mixed impact</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Positive impact</td>
<td>2</td>
<td>42</td>
<td>8</td>
</tr>
</tbody>
</table>

### Table 3: Visual arts (n=26)

<table>
<thead>
<tr>
<th></th>
<th>Not determined</th>
<th>Weak</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>No beneficial impact</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Mixed impact</td>
<td>8</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Positive impact</td>
<td></td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Dance (n = 15)

<table>
<thead>
<tr>
<th></th>
<th>Not determined</th>
<th>Weak</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>No beneficial impact</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Mixed impact</td>
<td>3</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Positive impact</td>
<td></td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
### Table 5: Drama (n = 27)

<table>
<thead>
<tr>
<th></th>
<th>Not determined</th>
<th>Weak</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>No beneficial impact</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Mixed impact</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Positive impact</td>
<td>13</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>suggested</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 6: Multi-arts (n = 38)

<table>
<thead>
<tr>
<th></th>
<th>Not determined</th>
<th>Weak</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>No beneficial impact</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Mixed impact</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Positive impact</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>suggested</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>