Welch, G.F., Himonides, E., Saunders, J., Papageorgi, I., Rinta, T., Preti, C., Stewart, C., Lani, J., & Hill, J. (in press). Researching the first year of the National Singing Programme in England: an initial impact evaluation. *Psychmusicology: Music Mind and Brain.* [Special Issue on Singing]

Researching the first year of the National Singing Programme *Sing Up* in England: an initial impact evaluation

Welch, G.F., Himonides, E., Saunders, J., Papageorgi, I., Rinta, T., Preti, C., Stewart, C., Lani, J., and Hill, J.

Institute of Education, University of London, UK

Abstract

The article reports on the first year of an independent evaluation of the National Singing Programme Sing Up (2007-2011) in England. The aim of the programme is that, by 2012, Primary-aged children (up to the age of 11 years) should be experiencing high-quality singing, both within and outside their daily school curriculum, on a daily basis. As part of the evaluation, eighty-one schools were visited across England and n=3,762 individual children had their singing behaviour and development assessed using a combination of two established rating scales. The focus was on providing an initial profile of children's singing prior to the launch of the National Programme and, subsequently later in the first year, to conduct a small number of additional visits to see if there was any evidence of early impact. The emergent data confirm that children's singing is subject to developmental processes, with variations related to sex, age and ethnicity. There is also some evidence that, notwithstanding such variations, a programme of sustained singing education can have a positive benefit on children's singing behaviours and development.

Introduction and background

Research suggests that singing behaviours are subject to developmental processes in which individual neuropsychobiological potentiality is shaped (nurtured and/or hindered) by learning experiences within socio-cultural contexts (Welch, 2007; in press; Knight, 2009). Although singing is commonplace, it is also marked by cultural diversity, with development related to opportunity (e.g., Mang, 2007), the prosodic features of

indigenous languages (Azechi, 2008), as well as the dominant characteristics of the local musical soundscapes (Welch*,* Sergeant, & White, 1997; Welch, 2006a; 2006b; in press).

In many parts of the world, the ability to sing is seen as a mark of an individual's underlying musicality (*cf* Sloboda, Wise, & Peretz, 2005). Consequently, those individuals whose singing development has been hindered in some way are often labelled (including self-labelled) in some absolutist sense under a bi-polar categorisation of 'can'/'cannot' sing, with variations in their ascribed musical identity as a 'non-singer', 'tone-deaf', or 'tone-dumb' being found in virtually all cultures. Yet, as mentioned above, contrary evidence from developmental and neurological studies continues to emerge that singing and musical behaviours are context bound and susceptible to improvement with appropriate experience which can be informal as well as formal (e.g., Brown, Martinez, Hodges, Fox & Parsons, 2004; Koelsch, Fritz, Schulze, Alsop & Schlaug, 2005; Mang, 2006; 2007; Dalla Bella Guguère & Peretz 2007; Kleber, Veit, Birbaumer & Lotz, 2007; Fuchs, Meuret, Thiel, Täschner, Dietz & Gelbrich, 2007; Mithen & Parsons, 2008; Stewart & Williamon, 2008; Welch*,* Himonides, Saunders, Papageorgi, Rinta, Stewart, Preti & Lani, 2008; see Welch, 2006a for review).

Furthermore, the recent wealth of studies into the neurosciences and music (*cf* Avanzini, Faienza, Minciacchi, Lopez & Majno, 2003; Avanzini, Koelsch, Lopez & Majno, 2005) continue to amass evidence of the multi-sited representation of musical behaviours in various regions of the brain, including singing (Kleber *et al*, 2007). These and related studies also indicate that there are various other-than-musical benefits that can accrue for the individual from engaging in musical (including singing) activity, such as related to physical and psychological health and well-being (Clift & Hancox, 2001; Clift, Hancox, Morrison, Hess, Kreutz & Stewart, 2007; Kreutz, Bongard, Rohrmann, Hodapp & Grebe, 2004; Welch, 2005), social skill development and social inclusion (Odena, 2007; Portowitz, Lichtenstien, Egorov & Brand, 2008) and cognitive development (Schlaug, Norton, Overy & Winner, 2005).

Therefore, it is perhaps not surprising that music is figuring more significantly in the contemporary educational policies of several of the world's governments. Brazil, for example, has recently (August, 2008) signed into law its first compulsory provision of music education in schools; Italy's regional government for Emilia Romagna has supported a special project (2005-2008) concerning the provision of vocal and choral education in primary schools; Poland has a *Sing Poland* programme for the development of choral singing in schools; and the UK Government has a formal 'Music Manifesto',

defined by the (then) Department of Culture, Media and Sport (DCMS) as '…a campaign for improvement in music education. It is about creating more music for more people.'

The Music Manifesto was launched in July 2004 by the (then) Schools Minister, David Miliband, and the Arts Minister, Estelle Morris, together with 60 plus signatories from the music industry. It is described on the Music Manifesto website as

> '…the result of a unique collaboration between the DfES and DCMS [two Government Ministries] with music organisations, musicians, teachers and composers, the music industry, broadcasting, teacher and musicians' unions, arts and education charities and Trusts' (see http://www.musicmanifesto.co.uk/history [retrieved 21 July 2007]).
>
> 'At the heart of the Music Manifesto is a desire to see more opportunities in music for more young people – from high quality curriculum tuition to out of school hours youth and garage bands; from composing to live performance, from classical concerts to DJing and gigs. In its final form, the Music Manifesto offers a strategic direction for the future of music education and a common agenda for joint action' (ibid).

The Department for Children, Schools and Families (DCSF – created in 2007) reported that its predecessor, the Department for Education and Skills (DfES) had invested over £500 million in music education between 1999 and 2008, with £95 million proposed for investment in 2007/08 alone.

One major component of the 'Music Manifesto' concerns the *National Singing Programme* (2007-2011). This arose from a recommendation in the 2nd Report of the Music Manifesto group ('Making every child's music matter' October 2006) that singing be provided for all early years and primary-aged children by 2012.

> 'Singing offers the most direct route to providing a music-making experience for all children and young people, so we believe it should be a central element of the universal music offer. As a result, we recommend putting group singing at the heart of all primary school musical activity.' (Music Manifesto Report No 2, 2006:8)

In part, this was because of the opportunity afforded for the development of a cultural programme (2008-2012) that would be linked to the 2012 London-based Olympic Games (Education Guardian, 18th October, 2006). But it was also in recognition of the perceived importance of singing as a foundation for all round music education development. In the introduction to the 2nd Music Manifesto report, Marc Jaffrey, the 'Music Manifesto Champion' wrote:

> 'Singing has the potential to involve children and young people in music on a scale that we have not witnessed before. It is the most elemental form of music

> making, and is within the grasp of all of us, whatever our ability. It is a powerful community activity binding individuals and community together.'

The UK Government's then Secretary of State for Education and Skills, Alan Johnson, together with the then Culture Minister, David Lammy, announced in January 2007 the launch of an additional £10m funding package (subsequently confirmed as £10m per year for four years) to support school singing, both in and out of school hours, through a major national singing campaign for primary school aged children, led by the British composer and broadcaster Howard Goodall in a new role as the 'Singing Ambassador' for England (DfES Press Notice, 16th January 2007 - http://www.dfes.gov.uk/pns/DisplayPN.cgi?pn_id=2007_0009).

Following a tendering process, the two Government Departments (DCMS, DfES) jointly appointed a consortium of Youth Music, The Sage Gateshead, Faber Music, and advertising agency Abbot Mead Vickers to lead on the actual provision of the National Singing Programme in 2007-2008 and on through to 2011. Included in the intentions of the Programme are that 'children experience high-quality singing, both within and without their daily school curriculum, on a daily basis' and that 'Every school has a teacher committed to facilitating high quality singing and vocal work for the whole school'.

The *Sing Up* National Singing Programme was launched in November 2007 and a team from the Institute of Education, University of London, led by the first author, was appointed to undertake a research evaluation of key elements of the Programme. Two prime foci were: (i) to undertake an initial 'baseline' audit of children's singing behaviours in randomly selected schools and (ii) to link this 'baseline' data collection to a post-impact evaluation of particular *Sing Up* programme interventions with children and adults (teacher, parents and other professionals involved in promoting singing in community contexts). This paper focuses on the evidence concerning the possible impact of the programme on participant children's singing behaviours during the first year (to July 2008).

Methodology

Participants

The participants for the first year's (2007-2008) research ('baseline' and post-intervention evaluation) were drawn from 81 schools located across England. The schools were in major cities and adjacent population centres across the South-East (Greater London),

East (Cambridgeshire, Essex), South-West (Bristol and Gloucester), Midlands (Birmingham, Coventry, Derby), North-West (Manchester) and North-East of England (Newcastle, Gateshead, Durham, York), supplemented by a smaller number of schools in other parts of the country in urban, suburban and rural settings, as well as a group of Cathedral Choir Schools. Contacts were made initially with Local Authority music advisors and university music education colleagues for advice on possible participant schools (see acknowledgements at the end of this article), the intention being to draw on local knowledge to ensure that a diverse range of school singing 'cultures' were accessed, i.e., schools with a known history of good singing and those without. Cathedral Choir Schools were contacted directly.

Within each school, participant children were drawn mainly from two contrasting age groups, 7-year-olds and 10-year-olds, representing the upper and lower age groups within Key Stage 2 of the National Curriculum in England. Previous research (e.g., Rutkowski, 1997; Stadler Elmer, 2002; Welch, 1998; 2006a; 2006b; 2007) had demonstrated that clear developmental differences in singing behaviour by age and sex were likely to be evidenced by the selection of these two age groups. Other recent findings from research into the acoustics of children's singing voices (Sergeant & Welch, 2008; 2009) and children's vocal health in singing and speaking (Rinta & Welch, 2008; Williams, Welch & Howard, 2005) similarly supported such a conception. However, where the prime age focus 'baseline' children were in classes with mixed age groups (such as 10-year-olds with some 9-year-olds), normally all the children in the class were assessed in order to ensure that none felt excluded. This meant that the 'baseline' data also included some 6, 8 and 9-year-olds, as did the classes assessed in the post-intervention data collection.

All participants (headteachers, teachers and pupils) had the purpose of the assessment explained in advance (and in writing to the school). Under our ethical guidelines, we guaranteed anonymity to all participants and reminded them that they were allowed to withdraw from the assessment process at any time that they felt uncomfortable.

The initial assessment phase ran from late September 2007 through to February 2008 and was focused on generating some sense of the commonality and diversity of singing behaviours across pupils in English Primary schools. This phase was termed the Year 1 "baseline' assessment. Subsequently, these "baseline' assessment' schools formed part of

a larger group of 'Non-Sing Up' schools as the research unfolded across successive years. These schools (of either "baseline"/Non-Sing Up label) were so-labelled because they were not participating in the Sing Up national programme at the time of our assessment.

In this 'baseline' phase, n=3,510 children were assessed from 77 schools (see below for the assessment protocol). Of these, 10 schools subsequently were visited again, that is, they had one visit during the 'baseline' phase and then again between May and July 2008 after a specific *Sing Up* singing development focused intervention. These 10 schools accounted for n=495 assessments within the 'baseline' phase and n=324 assessments post-intervention (see Table 1a). The post-intervention number of assessments was smaller because not all of the original 'baseline' children were involved subsequently in the intervention (because the classes selected for the intervention were at the school's discretion). In addition, another n=70 children from 4 schools were assessed only after their *Sing Up* intervention, but not before. Together, this makes a total of 14 schools and 394 assessments in the post-intervention phase. The breakdown of numbers of participants by age and sex for each assessment phase is shown in Table 1b. The total numbers of individual pupils involved across all assessment phases was 3,762.

Table 1a: Numbers of individual participant assessments by school and phase ('baseline' n=3,510; and post-intervention n=394 [324+70])

	'baseline' Phase	**Post-intervention Phase**	Grand Total
School visited twice	495	324	819
School visited during 'baseline' phase only	3015	-	3015
School visited during post-intervention phase only	-	70	70
Grand Total	*3510*	*394*	*3904*

Table 1b: Numbers of participants by age and sex for each phase ('baseline' n=3,510; and post-intervention n=394 [324+70])

The amount of time between a 'baseline' assessment and a follow-up post-intervention assessment was within a maximum period of nine months within the schools year 2007-2008, i.e., from October 2007 to the following June 2008. Although there may be differences in the data that relate to variations within this assessment process, the current data analyses have assumed that the 'baseline' to post-intervention

assessment period is relatively similar for the n=324 children involved, being approximately six to eight months. As more data is collected it will be possible to take a more exact measure by looking at children's dates of birth and plotting these against the assessment data.

	Boys	**Girls**	Grand Total
Year 2	77	68	145
Year 3	769	703	1472
Year 4	55	56	111
Year 5	170	114	284
Year 6	780	718	1498
'baseline' totals	*1851*	*1659*	*3510*
Year 2	41	39	80
Year 3	33	46	79
Year 4	41	44	85
Year 5	38	54	92
Year 6	25	33	58
Post-intervention totals	*178*	*216*	*394*
Grand Total	*2029*	*1875*	*3904*

Notwithstanding the essential school class-based organisation of the 'baseline' study and the *Sing Up* interventions, as each child had been assigned a unique reference number within the database, it was possible to identify subsequently those individual children (n=107) who had been assessed during the 'baseline' phase and again after they had their *Sing Up* input.

Assessment Protocol

As mentioned above, amongst the prime foci of the research evaluation of *Sing Up* in its first year (2007-2008) were to create an initial 'baseline' profile of (a) children's singing and vocal behaviours and (b) attitudes to singing that could be used for comparative purposes, i.e., during the research visit, class teachers made provision for each child to complete a 45-question survey of their attitudes to singing. This has been reported elsewhere, as the data analyses (including factor analyses) are extensive and beyond the space available in this paper – see Welch, Himonides, Saunders, Papageorgi, Rinta, Preti,

Stewart, Lani, Vraka & Hill (2008) (http://imerc.org/research_nsp.php). Another focus was to evaluate the possible impact of a specific *Sing Up* intervention on participant children in this early phase of the programme's enactment. This particular intervention was termed 'Singing Playgrounds' and was provided by members of *Ex Cathedra*, one of the UK's leading choir and Early Music ensembles. 'Singing Playgrounds' is an educational outreach programme designed to develop children's musicianship through singing games. Expert adult singers visit school playgrounds and work with older children – usually called 'Song Leaders' – who lead their peers in singing games.

> 'Through the use of weekly set tasks, the Song Leaders are encouraged to develop and evaluate their own activities. Equipped with clipboards and stickers to hand out to the younger children for enthusiastic participation in "Jump Jim Joe" and other popular playground hits, the song leaders…are seen as role models throughout the school and are chosen for their enthusiasm.' (retrieved 18 August 2008 from http://www.singup.org/teachers_and_music_leaders/recipes_for_success/Singing_Playgrounds.php)

The research protocol for the assessment of singing and other vocal behaviours drew on established models on singing behaviour and development from the literature. Previous research indicated that it would be helpful to assess more than one aspect of children's vocal behaviour in order to build a composite, rounded picture. Consequently, the protocol investigated: (i) the children's *habitual speech pitch centre* (by asking each participant to count backwards from either ten or twenty – depending on age – and noting the spoken pitch centre in relation to an adjacent piano keyboard, n.b., the process of counting backwards is reported to be close in terms of fundamental frequency (F_0) to conventional speech, whereas normal counting is more declarative in manner and believed to be more akin to reading, at least in adults, which is likely to have a higher average spoken frequency (Johns-Lewis, 1986)); (ii) *comfortable singing range* (by imitative singing of a musical song fragment at various starting pitches, transposed upwards and downwards with reference to an adjacent keyboard; comfortable singing range, rather than singing range limits, is regarded as a more valid measure of children's customary singing behaviour with regard to song items in their local culture (Welch, 1979)); and (iii) *singing behaviour of two well-known song items* (either 'Twinkle, Twinkle' and 'Happy Birthday' or, if these were unknown – on advice from the teacher – one or other items that the particular child knew well). The last of these three elements was assessed against two established rating scales (Rutkowski, 1997; Welch, 1998) (see Figure 1).

Rutkowski (1997) *Singing Voice Development Measure (SVDM)*

1	"Pre-singer" does not sing but chants the song text.
1.5	"Inconsistent Speaking Range Singer" sometimes chants, sometimes sustains tones and exhibits some sensitivity to pitch, but remains in the speaking voice range (usually a3 to c4 [note: the pitch labels have been altered to bring them in line with modern conventions in which middle C = c4, 256 Hz]).
2	"Speaking Range Singer" sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range (usually a3 to c4).
2.5	"Inconsistent Limited Range singer" waivers between speaking and singing voices and uses a limited range when in singing voice (usually up to f4).
3	"Limited Range Singer" exhibits consistent use of initial singing range (usually d4 to f4).
3.5	"Inconsistent Initial Range Singer" sometimes only exhibits use of limited singing range, but other times exhibits use of initial singing range (usually d4 to a4).
4	"Initial Range Singer' exhibits consistent use of initial singing range (usually d4 to a4).
4.5	"Inconsistent Singer" sometimes only exhibits use of initial singing range, but other times exhibits use of extended singing range (sings beyond the register lift: b^b4 and above).
5	"Singer" exhibits use of extended singing range (sings beyond the register lift: b^b4 and above).

Welch (1998) *A revised model of vocal pitch-matching development (VPMD)*

Phase 1	The words of the song appear to be the initial centre of interest rather than the melody, singing is often described as 'chant-like', employing a restricted pitch range and melodic phrases. In infant vocal pitch exploration, descending patterns predominate.
Phase 2	There is a growing awareness that vocal pitch can be a conscious process and that changes in vocal pitch are controllable. Sung melodic outline begins to follow the general (macro) contours of the target melody or key constituent phrases. Tonality is essentially phrase based. Self-invented and 'schematic' songs 'borrow' elements from the child's musical culture. Vocal pitch range used in 'song' singing expands.
Phrase 3	Melodic shape and intervals are mostly accurate, but some changes in tonality may occur, perhaps linked to inappropriate register usage. Overall, however, the number of different reference pitches is much reduced.
Phase 4	No significant melodic or pitch errors in relation to relatively simple songs from the singer's musical culture.

Figure 1: Two independent measures of singing behaviour and development, each used in the assessment of participants' singing behaviours on two selected song items

Previous research (Mang, 2006) had demonstrated that the two scales could be used alongside each other to investigate complimentary aspects of singing behaviour and development. The Rutkowski (1997) scale is a measure of singing voice development, whereas the Welch (1998) scale assesses vocal pitch-matching development.

Children were visited at their schools where their singing and vocal behaviours were assessed individually in a quiet space. Each child was taken through the assessment protocol, normally being tested individually within a small group of between two to four children that was drawn from the class. This allowed the other members of the group to observe and see what was required, as this had been shown previously to be an appropriate method of accessing better quality responses than individual testing alone (*cf*

Plumridge, 1972). Children tended to be less nervous and, if shy, able to understand more clearly what was expected of them by listening to their peers in advance. To avoid the effects of vocal modelling, no starting pitch was given for the song items and, although the member of the research team provided verbal encouragement to the child, they did not offer any sung prompt (*cf* as advised by Mang, 2006). All children completed the assessments and none was excluded from the study.

Because of the large numbers of participants it was necessary to create a relatively large research team to undertake the fieldwork. Consequently, to promote reliability in the assessment process, this was undertaken initially by moderation, with members of the research team undergoing initial training on sampled items, then undertaking a school visit in pairs prior to making visits on their own. The validity and ease of use of the assessment protocol was established through a short piloting process prior to commencement of the main data collection.

The piloting process involved two members of the research team visiting a local Primary school and using the draft protocol to audio record digitally individual children of different ages. The resultant vocal products were then put on line, duplicated and randomised and then rated by both themselves and other individual members of the team according to the two assessment scales. The rating results were compared statistically and revealed a close agreement amongst the individual members of the team (Kendall's Coefficient of Concordance, $W(5,19)=.909$, $p=.000$). The pilot sound files and ratings were reviewed and discussed subsequently at a team meeting prior to the initial 'baseline' visits being undertaken and moderation of initial visit ratings prior to the main programme of individual researcher visits. An accumulated plotting of the means from the ratings of each member of the research team after their visits revealed that only one judge had mean data for one school that was an obvious outlier. This was reviewed and adjusted after internal discussion and clarification.

Participants' responses were noted onto individual assessment forms (see Welch *et al*, 2008 for an example) and data were subsequently entered for collation and analysis into a specially designed data entry form that was connected to a structured query language (SQL) based database (SQL is a database computer language designed for managing data in relational databases, e.g., see http://en.wikipedia.org/wiki/SQL). Each participant was uniquely coded in order to enable comparative assessment of their singing behaviour and development at a later date as necessary. The database included information on participant demographics (research sites, child's age, year group, sex,

ethnicity), unique identification codes for each child, spoken pitch centre, ratings on each of the two independent measures of singing behaviour and development and normalised singing score, being a conversion of the rated measures into a percentage of the maximum ratings across the combined rating scales. If a child were rated at the highest possible level on each of the two independent measures for both song items, this would equate to 100% normalised singing score. Slightly lower ratings generate lower percentages and equivalent normalised singing scores. This data processing facilitates the possibility of group comparison, such as by class, school, age, sex or ethnicity.

Results

Prime foci for this paper were (a) to gain insight into the nature of children's singing behaviour and development and (b) to explore the evidence of any early *Sing Up* programme impact arising from the 'Singing Playgrounds' intervention on the singing behaviours of participant children.

A univariate Analysis of Variance (ANOVA) investigated any possible main and interaction effects between the four main participant factors (sex $^{\text{male, female}}$; ethnicity $^{\text{Asian, Black, White, Chinese, Mixed, Other}}$; school year group $^{\text{Year 2, 3, 4, 5, 6}}$; intervention $^{\text{'baseline', post-intervention}}$) on the observed normalised singing ratings. The analysis suggested that there were significant differences evidenced overall, $F(87, 3903) = 6.142$, $p=.000$ (Levene's Test for Equality of Error Variances applied), including for sex $F(1,3903) = 8.108, p<.005$, ethnicity $F(1,3903) = 4.55, p<.05$, school year group $F(4,3903) = 3.141, p<.002$ and intervention $F(5,3903) = 2.606, p<.05$ - see Table 2). There were three interaction effects evidenced. These were for sex and year group $F(4,3903) = 2.595$ $p<.05$, ethnicity and year group $F(17,3903) = 2.272, p<.005$ and for intervention and year group $F(4,3903) = 2.991, p<.005$ (see Table 2).

Table 2: Tests of main and interaction effects between the three main participant factors of sex, ethnicity and school year group

Tests of Between-Subjects Effects

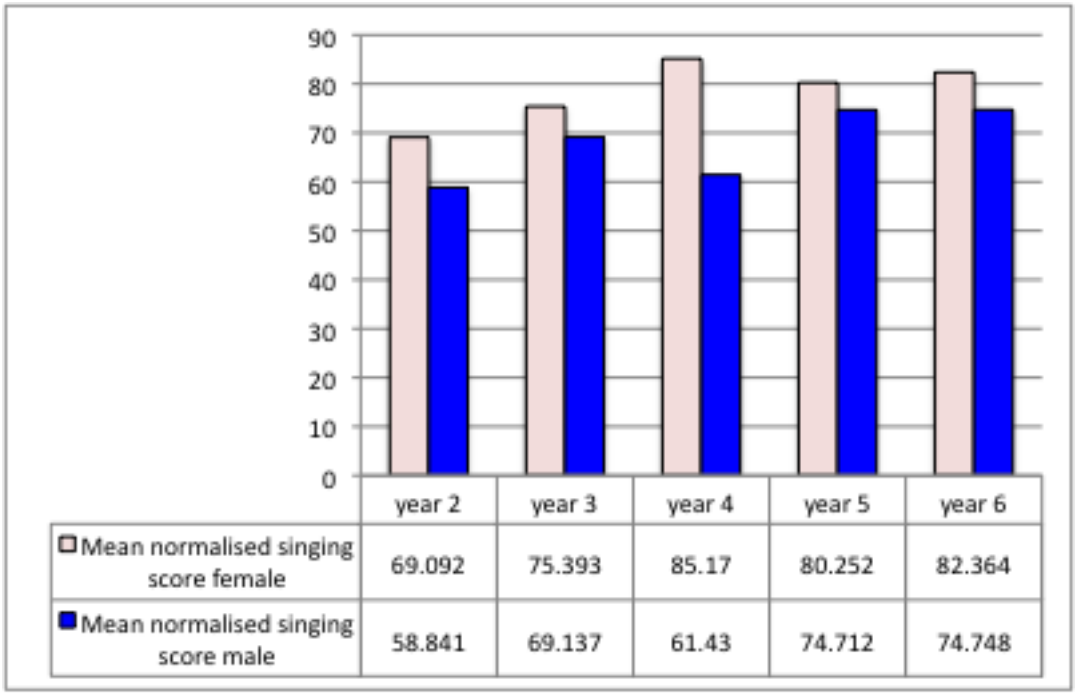
Dependent Variable: Normalised Singing Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.051E5	87	2357.839	6.142	.000
Intercept	834310.269	1	834310.269	2173.458	.000
Sex	3112.509	1	3112.509	8.108	.004
Ethnicity	5001.254	5	1000.251	2.606	.023
Tempus ('baseline'/post-intervention)	1746.493	1	1746.493	4.550	.033
Yeargroup	4822.554	4	1205.638	3.141	.014
Sex * Ethnicity	1131.660	5	226.332	.590	.708
Sex * Tempus	20.143	1	20.143	.052	.819
Sex * Yeargroup	3984.540	4	996.135	2.595	.035
Ethnicity * Tempus	3044.620	4	761.155	1.983	.094
Ethnicity * Yeargroup	14829.561	17	872.327	2.272	.002
Tempus * Yeargroup	4592.753	4	1148.188	2.991	.018
Sex * Ethnicity * Tempus	225.295	3	75.098	.196	.899
Sex * Ethnicity * Yeargroup	4665.793	14	333.271	.868	.594
Sex * Tempus * Yeargroup	450.848	4	112.712	.294	.882
Ethnicity * Tempus * Yeargroup	6238.187	11	567.108	1.477	.133
Sex * Ethnicity * Tempus * Yeargroup	987.277	8	123.410	.321	.958
Error	1.465E6	3816	383.863		
Total	2.105E7	3904			
Corrected Total	1.670E6	3903			

(i) Independent measures of singing behaviour and development by age, sex and phase of assessment

There is evidence of sex differences in singing behaviour overall as reported in the ANOVA. Overall, across the whole dataset, boys (M=68.873, n=2029,) tended to have a significantly lower mean rating than girls (M=79.103, n=1875, $t(3900)=11.2$, p=.000). There is also evidence of sex differences in relation to children's ages. In general, older children (age 10+, School Year 6) tend to be rated as more developed than their younger peers (e.g. age 7+, School Year 3) and girls tend to be rated as more developed than boys

in each age group. This is evidenced on both rating scales (Rutkowski; and Welch) separately and also in the combined normalised score for each age group (see Figure 2). There is a significant correlation between the singing behaviour assessments using the two ratings scales (Rutkowski, Welch) across the two focus songs, varying between r=.824 and .939, all correlations at p<.001.



	year 2	year 3	year 4	year 5	year 6
Mean normalised singing score female	69.092	75.393	85.17	80.252	82.364
Mean normalised singing score male	58.841	69.137	61.43	74.712	74.748

Figure 2: Overall normalised singing scores by sex and age (school year group)

These figures were broken down to investigate the 'baseline' and post-intervention datasets and a series of t-tests was undertaken to compare the means for the sexes within each age group (Tables 3 and 4). These indicate that there is no difference between the sexes in the 'baseline' data for Year 2 (age 6), but that for each successively older age group (with two exceptions), the girls tend to be rated more highly than the boys. The exceptions are in Year 5 in the 'baseline' data and Year 6 in the post-intervention scores. (Note: these analyses for Years 2, 4 and 5 which drew on data from relatively small numbers of children became more robust in subsequent assessment years when they became the prime age foci for extending the scope of the year one study reported here. Nevertheless, overall, the post-intervention assessments for the year one research indicated that both boys and girls were generally having their singing behaviour

being rated as more advanced developmentally following their 'Singing Playgrounds' experiences.)

Table 3: Mean normalised singing scores by sex, school year group and 'baseline' versus post-intervention

Dependent Variable: Normalised Singing Score

Sex	Research tempus	Yeargroup	Mean	Std. Error
female	'baseline'	Year 2	56.191	6.753
		Year 3	70.205	1.919
		Year 4	86.267	7.214
		Year 5	78.944	3.760
		Year 6	76.325	1.943
	post intervention	Year 2	81.992	6.936
		Year 3	83.175	3.649
		Year 4	84.511	3.910
		Year 5	82.494	3.072
		Year 6	89.611	6.032
male	'baseline'	Year 2	53.067	7.048
		Year 3	64.629	2.643
		Year 4	66.546	7.245
		Year 5	74.405	4.384
		Year 6	68.306	2.228
	post intervention	Year 2	70.388	3.696
		Year 3	75.899	6.264
		Year 4	56.314	5.554
		Year 5	75.080	5.211
		Year 6	87.633	5.363

Table 4: Normalised comparative ratings for year group and sex in each assessment phase and combined (complete dataset)

	male versus female		
	'baseline'	Post-intervention	Complete dataset
Year-2	N.S.	N.S.	N.S.
Year-3	-6.382***	-2.877**	-6.931***
Year-4	-4.720***	-4.742***	-6.690***
Year-5	N.S.	-2.573*	-2.757*
Year-6	-6.931***	N.S.	-7.081***
	*p<0.05 **p<0.005 ***p<0.0001		

(ii) Additional evidence of impact

An initial independent samples t-test was conducted in order to compare the normalised singing competency scores (Rutkowski & Welch combined ratings) for all pupils that were assessed during the 'baseline' (n=3510) with post-intervention periods (n=394) in year one. The difference between the two (equal variances not assumed) was statistically significant [$t(539)=11.2$, p<.0005]. The overall scores for the 'baseline' (M=69.425, SD=20.825) were significantly lower that those for the post-intervention data (M=79.714, SD=16.781). Although the magnitude of the difference of the means was small (eta squared=0.031), there is a markedly higher score in assessed singing competency for the post-intervention pupils.

Similarly, an independent samples t-test was conducted in order to compare the normalised singing competency scores of pupils whose schools had been visited both during the 'baseline' (n=495) and post-intervention (n=324) phases. There was a significant difference [$t(762)=11$, p<.0005, equal variances not assumed] between 'baseline' assessments (Visit 1) (M=65.72, SD=19.792) and post-intervention assessments (M=79.96, SD=16.898). The magnitude of the difference of the means was quite large (eta squared=0.129).

Furthermore, within the ten schools that the team visited twice, the number of individual pupils that were assessed during both visits and whose data can be matched was n=107. Accordingly, a paired samples t-test was run on their normalised scores. This revealed a statistically significant improvement [$t(106)=5.916$, p<.0005] between the 'baseline' phase assessments (M=70.58, SD=16.096) and the post-intervention assessments (M=81.80, SD=15.355). The eta squared (0.25) indicates a large size effect.

(iii) Evidence of impact on sung vocal pitch ranges

A further analysis was undertaken to compare the comfortable sung pitch ranges of the same n=107 children. A paired-samples t-test was applied to evaluate the impact of the *Sing Up* intervention on each individual's comfortable sung pitch range in semitones. There was a statistically significant increase [$t(106)=5.398$, $p<0.0005$] between the 'baseline' phase range (M=10.83, SD=5.614) and sung pitch range produced in the post-intervention phase (M=13.70, SD, 4.379) of three semitones. The eta squared statistic (0.22) indicated a large size effect. At this stage in the research process, it was not known whether this improvement in comfortable sung pitch range was due to the specific intervention, or due to a normal process of maturation. However, it is possible to make a cautious inference that, as their singing development rating had improved, there may also be a concomitant improvement in comfortable sung pitch range. (Subsequent analyses of data from the second year of data collection indicate that this comfortable singing range increases with age extend from g3 to c5 at age 7+ (approximately one and a half octaves) through to f3 to e ♭5 at age 10+ (almost two octaves). (Welch *et al*, 2009).)

(iv) Evidence of impact related to participant ethnicity

Schools provided background information on the ethnicity of pupils according to the Department for Children, Schools and Families official classification. An analysis of normalised singing scores for each of these groups indicates that, within and across the two data collection phases, Asian pupils (n=609; M=65.38, SD=21.217) tended to score significantly lower than their White peers (n=2729; M=71.15, SD=20.308) [$t(874)=6.12$, p.=.000, equal variances not assumed]. Asian pupils also scored significantly lower than their Black peers (n=294; M=72.46, SD=21.458) [$t(874)=6.12$, p.=.000, equal variances not assumed]. There were no significant differences between White and Black children in either 'baseline' or post-intervention data. (The numbers of pupils in the other ethnicity categories in the Year 1 data were too small for meaningful statistical comparison.)

However, notwithstanding these statistical differences between ethnicities, all three major groups have significantly higher normalised singing ratings in their post-intervention assessment data (Asian pupils $t(253)=7.078$, $p<.0005$; Black pupils $t(113)$-5.414, $p<.0005$; White pupils $t(221)=10.478$, $p<.0005$; equal variances not assumed). In the case of Asian pupils, the post-intervention score is also much higher than that for the White and Black pupils at 'baseline' (see Figure 3 and Table 5).

Table 5: Comparison of 'baseline' and post-intervention data by ethnicity and normalised singing ratings (using Official UK Government ethnicity categories)

ethnicity	'baseline' N	normalised 'baseline' singing score	post-intervention N	normalised post-intervention singing score	Grand Total N	Grand Total singing score
Other	86	70.7	0	0.0	86	70.7
Asian	462	62.1	147	75.6	609	65.4
Black	239	69.9	55	83.4	294	72.5
Chinese	14	74.0	5	79.0	19	75.3
Mixed	149	72.8	18	79.0	167	73.5
White	2560	70.4	169	82.2	2729	71.2
Grand Total	3510	69.4	394	79.7	3904	70.5

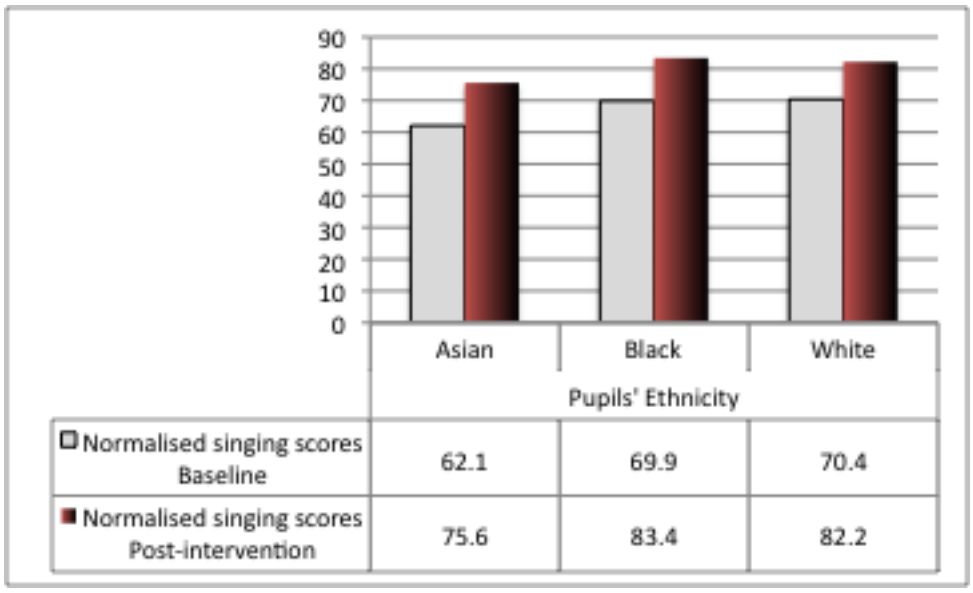


Figure 3: Mean normalised singing scores for 'baseline' and post-intervention ratings by ethnicity for the three main groupings (Asian, Black, White)

Discussion

Some caution is needed in the interpretation of the statistical data concerning evidence of impact for several reasons. Firstly, it may be that schools that are choosing to participate in the national programme already have a senior management team that is committed to fostering singing within the curriculum. We tried to counter this by ensuring that our 'baseline' data embraced many different levels of commitment to, and engagement with, singing. Nevertheless, there may be an unintended bias in the data because our post-intervention schools had voluntarily agreed to take up an offer to participate in the

national programme from the organisers and had invited the *Ex Cathedra* musicians to enrich the music curriculum through their 'Singing Playgrounds' activities. It has been necessary to keep this possible bias under review in subsequent data collection and analyses in successive years. Secondly, caution is necessary because the data set is (of necessity) uneven in its distribution across age groups, and the numbers of pupils that were assessed in the post-intervention phase was somewhat smaller than that available for the overall 'baseline'. It is also clear from the general trend in the data and earlier literature (e.g. Mang, 2006; Rutkowski, 1997; Welch *et al*, 1997) that children are likely to improve in their singing behaviours as they get older, including within the maximum of nine months embraced by these two phases of data collection. Nevertheless, there are some positive early indications that at least one major strand of the *Sing Up* National Singing Programme – the 'Singing Playgrounds' initiative by *Ex Cathedra* – is making a difference to the underlying singing behaviours of the participant children. Taken together with survey evidence from the same children's reported attitudes to singing (Welch *et al*, 2008) – that also shows a positive shift towards maintaining engagement with singing in both boys and girls – the initial impression is of an upward trajectory in these participants' singing engagement and development during this launch year likely because of the new national programme. As suggested above, where differences in children's normalised singing behaviours are seen to exist at school level in the 'baseline' data, anecdotal evidence (explored more systematically in the coming months) suggests that this relates to school leadership decisions on the relative importance of singing and music in the curriculum compared to other subjects.

As researchers, our prime aims were to establish some form of initial 'baseline' that could act as a comparison data set against which any post-intervention assessments could be measured. Although the numbers of children with data in the post-intervention phase were relatively small within the overall total reported here (in part because the *Sing Up* programme did not begin to roll out in schools until late 2007/early 2008), the basic assessment framework appeared to be robust and provided useful data on which an independent evaluative assessment of impact could begin to be made. Those children who were tracked from before and after their 'Singing Playgrounds' activities demonstrated a positive development in their singing abilities. Although numbers are small (n=107), they are part of a larger group (n=394) who also demonstrate more advanced singing behaviour developmentally.

The ambition of the UK Government is to develop singing for over 3 million children aged 5 to 10 years across 17,504 Primary schools and other community settings in England over a four-year period (2007-2011). The challenge is enormous, but the early research data provide encouraging evidence about the possible success of the initial steps on this journey. The emergent data also provide evidence that the development of singing behaviours should be considered as a normal and achievable feature of children's musical engagement with the world around them, particularly when they are provided with new, positive singing experiences.

Postscript

Since the original drafting and subsequent refining of this paper for *Psychomusicology*, the research team have completed the second year of data collection (2008-2009) – see Welch *et al*, 2009 – and are currently (June, 2010) just finishing the third year (2009-2010). At this moment in time, according to the *Sing Up* organisers, approximately 87% of Primary schools in England are signed up to the programme. Also, over time, many different strands to the research have been included in addition to the foci reported here, such as supplementary investigations into other specific singing interventions, research into the impact on teachers and community musicians undergoing professional development as teachers of singing to children, an investigation of wider benefits from singing, and an ongoing field-based study into the nature of the teaching and learning processes that can impact positively (or negatively) on children's singing behaviour and development. Overall, the evidence for children's singing being subject to developmental processes as reported in this current article continues to be supported, alongside additional and more extensive complementary evidence that the programme has been able to impact positively on accelerating this development where children experience a rich and sustained singing education programme. For more information, please contact the first author.

Acknowledgements

The research team members wish to thank Maurice Walsh, Senior Vocal Tutor with Manchester Music Service; Ula Weber of *Ex Cathedra*; Dr Penelope Harnett, University of West of England; Dr Liz Mellor, York St John University; and Sarah Kekus and Edward Milner of the *Sage Gateshead* for their invaluable support in identifying participant schools. We are also extremely grateful to all the schools (pupils, teachers and

headteachers) for their time and commitment to participate in this national research activity.

References

Avanzini, G., Faienza, C., Minciacchi, D., Lopez, L., & Majno, M. (2003). (Eds.), *The Neurosciences and Music* (Vol. 999). New York: Annals of the New York.

Avanzini, G., Koelsch, S., Lopez, L., & Majno, M. (2005). (Eds.), *The Neurosciences and Music II.* (Vol. 1060). New York: Annals of the New York Academy of Sciences.

Azechi, N. (2008). Young Children's Rhythmic Behaviour in Singing: The Influence of Mother Tongue on Their Development. *Proceedings,* ICMPC10, Sapporo, Japan, 25-29 August 2008.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research 20*, 363 - 375.

Clift S., & Hancox G. (2001). The perceived benefits of singing: Findings from preliminary surveys of a university college choral society. *Journal of the Royal Society for the Promotion of Health, 121*, 248-256.

Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G., & Stewart, D. (2007). Choral singing and psychological wellbeing: Findings from English choirs in a crossnational survey using the WHOQOL-BREF. In A. Williamon & D. Coimbra (Eds), *Proceedings*, International Symposium on Performance Science, Porto, Portugal, 22-23 November, 2007, 201-207.

Dalla Bella, S., Giguère, J-F., & and Peretz, I. (2007). Singing proficiency in the general population. *J. Acoust. Soc. Am. 121* (2), 1182–1189.

Fuchs, M., Meuret, S., Thiel, S., Täschner, R., Dietz, A., & Gelbrich, G. (2007). Influence of Singing Activity, Age, and Sex on Voice Performance Parameters, on Subjects' Perception and Use of Their Voice in Childhood and Adolescence. *Journal of Voice,* [Published on line September 2007]

Johns-Lewis, C**.** (1986). Prosodic differentiation of discourse modes. In C. Johns-Lewis (Ed.), *Intonation in Discourse* (pp. 199–219). London: Croom Helm.

Kleber, B., Veit, R., Birbaumer, N., & Lotze, M. (2007). Neural correlates of professional classical singing. In A. Williamon & D. Coimbra (Eds). *Proceedings*, International Symposium on Performance Science, Porto, Portugal, 22-23 November, 2007, 335-343.

Knight, S. (2009). *An attributional study of adult 'non-singers': an analysis of ideographic, nomothetic and socio-cultural perspectives.* Unpublished PhD Thesis, Institute of Education, University of London.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *NeuroImage, 25*, 1068– 1076.

Kreutz G., Bongard S., Rohrmann S., Hodapp, V., & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol and emotional state. *Journal of Behavioral Medicine, 27*, 623-635.

Mang, E. (2006). The effects of age, gender and language on children's singing competency, *British Journal of Music Education, 23,* 161-174.

Mang, E. (2007). Effects of Musical Experience on Singing Achievement, *Bulletin of the Council for Research in Music Education*, 174, 75-92.

Mithen, S., & Parsons, L. (2008). Singing in the brain. *New Scientist,* 23 February, p28.

Music Manifesto http://www.musicmanifesto.co.uk/ (Retrieved 24 August 2008).

Odena, O. (2007). *Music as a way to address Social Inclusion and Respect for Diversity in early childhood.* Study Paper for the Bernard van Leer Foundation. Belfast: NFER at Queen's.

Plumridge, J.M. (1972). *The Range and Pitch Levels of Children's Voices, in relation to Published Material for Children's Voices.* Unpublished Diss. Dip. Adv. Study of Ed, University of Reading.

Portowitz, A., Lichtenstien, O., Egorov, L., & Brand, E. (2008). Underlying mechanisms linking music education and cognitive modifiability. In S. Malbran & G, Mota (Eds). *Proceedings.* 22nd International Seminar on Research in Music Education, Porto, Portugal, 13-18 July, 2008.

Rinta, T., & Welch, G.F. (2008). Should singing activities be included in speech and voice therapy for prepubertal children? *Journal of Voice. 22*(1), 100-112.

Rutkowski, J. (1997). The nature of children's singing voices: Characteristics and assessment. In: B.A. Roberts (Ed.), *The Phenomenon of Singing* (pp. 201-209). St. John's, NF: Memorial University Press.

Schlaug, G., Norton, A., Overy, K., & Winner, E., (2005). Effects of music training on the child's brain and cognitive development. *Ann. N.Y. Acad. Sci. 1060*: 219–230.

Sergeant, D.C., & Welch, G.F. (2008). Age-related changes in Long-Term Average Spectra of children's voices. *Journal of Voice*, *22*(6), 658-670.

Sergeant, D.C., & Welch, G.F. (2009). Gender differences in Long-Term-Average Spectra of children's Singing voices. *Journal of Voice, 23*(3), 319-336.

Sing Up http://www.singup.org/ (Retrieved 24 August 2008).

Sloboda, J. A., Wise, K. J., & Peretz, I. (2005). Quantifying tone deafness in the general population, *Ann. N.Y. Acad. Sci.* 1060, 255–261.

Stadler Elmer, S. (2002). *Kinder singen Lieder: Über den Prozess der Kultivierung des vokalen Ausdrucks.* Berlin: Waxmann.

Stewart, L., & Williamon, A. (2008). What are the implications of neuroscience for musical education? *Educational Research, 50* (2), 177-186.

Welch, G.F. (1979). Vocal range and poor pitch singing. *Psychology of Music*, *7*(2), 13-31.

Welch, G. F. (1998). Early childhood musical development. *Research Studies in Music Education, 11*, 27-41.

Welch, G.F. (2005). Singing as Communication. In: D. Miell, R. MacDonald, & D. Hargreaves (Eds.), *Musical Communication.* (pp. 239-259). New York: Oxford University Press.

Welch, G.F. (2006a). Singing and Vocal Development. In: G. McPherson (Ed.) *The Child as Musician: a handbook of musical development.* (pp. 311-329). New York: Oxford University Press.

Welch, G.F. (2006b). The musical development and education of young children. In: B. Spodek & O. Saracho (Eds.), *Handbook of Research on the Education of Young Children*. (pp. 251-267). Mahwah, N.J.: Lawrence Erlbaum Associates Inc.

Welch, G.F. (2007). Addressing the multifaceted nature of music education: an activity theory research perspective. *Research Studies in Music Education*, *28*, 23-38.

Welch, G.F. (in press). Culture and gender in a cathedral music context: An activity theory exploration. In M. Barrett (Ed.), *A Cultural Psychology of Music Education*. New York: Oxford University Press.

Welch, G.F., Sergeant, D.C. & White, P. (1997). Age, sex and vocal task as factors in singing 'in-tune' during the first years of schooling. *Bulletin of the Council for Research in Music Education, 133*, 153-160.

Welch, G.F., Himonides, E., Saunders, J., Papageorgi, I., Rinta, T., Stewart, C., Preti, C., & Lani, J. (2008). The *National Singing Programme* for Primary schools in England: An Initial 'baseline' Study. In W. Sims (Ed.). *Proceedings,* International Society for

Music Education 28th World Conference, Bologna, Italy, 20-25 July, 2008, 311-316.

Welch, G.F., Himonides, E., Saunders, J., Papageorgi, I., Vraka, M., Preti, C., & Stephens, C. (2009). *Researching the second year of the National Singing Programme in England: An ongoing impact evaluation of children's singing behaviour and identity.* Institute of Education, University of London. [pp95]. [ISBN 978-1-905351-10-7]

Williams, J., Welch, G.F. & Howard, D.M. (2005). An exploratory 'baseline' study of boy chorister vocal behaviour and development in an intensive professional context. *Logopedics Phoniatrics Vocology, 30*(3/4), 158-162.