Ear Playing and Aural Development in the Instrumental Lesson:
Results from a ‘Case-Control’ Experiment

David Baker
Institute of Education, University of London, UK

Lucy Green
Institute of Education, University of London, UK

Abstract
This article reports on a case-control experiment that was conducted in 2012 as part of the Ear Playing Project (EPP) at the Institute of Education, University of London. The EPP developed from the ‘informal learning’ strand of Musical Futures and engaged instrumental students in the UK in learning from specially-created audio recordings without notation. Sixteen matched pairs of instrumental students (age 10–14 years) were divided into a ‘control’ and ‘experimental’ group, with pupils taught using traditional instrumental teaching methods (without ear playing) in the former and with ear playing in the latter. The instrumentalists in both groups underwent aural tests prepared by the Associated Board of the Royal Schools of Music (ABRSM), which required played responses. Pre- and post-test scores were appraised for both groups for criteria including ‘Pitch’, ‘Rhythm’, ‘Contour’, ‘Tempo’, and ‘Closure’, as well as a holistic mark. With the ‘experimental’ students surpassing the ‘control’ group in every criterion assessed at the post-test stage, results suggest that playing by ear from a recording may be beneficial for children’s aural development.

Keywords
ear playing, informal learning, instrumental teaching, Musical Futures, playing by ear

Introduction
Research from the 1940s (Mainwaring, 1941) has proposed various benefits linked to ear playing for instrumentalists and instrumental students. These include its effect on sight reading (Luce, 1958; McPherson, 1993/1994; Sperti, 1970), musical memorisation (McPherson, 1993/1994), performance skills (Glenn, 1999; Haston, 2004) and wider aural
Priest (1988, 1989) argued that traditional instrumental teaching methods have meant that playing by ear has been undervalued, yet all playing can be viewed as ‘by ear’, even when notation is involved. He proposed a model based on imitation and invention. Priest’s work illustrated that musical reproductive and creative capacities may lie at the heart of all instrumental musicianship.

McPherson, Bailey and Sinclair (1997) designed and tested a model of performance types with 101 high school wind instrumentalists. The authors looked at various influences on the re-creative skill of performing a repertoire of rehearsed music and compared this with improvising. Sight reading, along with the length of time instrumentalists had been learning, was most influential on success at performing a repertoire, whereas improvising was most markedly affected by an ability to play by ear. Interestingly, these authors contended that, for those exposed to a traditional style of teaching, sight reading capability was linked to how well students played by ear. Their research highlighted the significance of ear playing in developing other realms of performance skill, including within conventional learning contexts. In these contexts, ear playing might typically be absent as teaching and learning focuses chiefly on print notation. McPherson and Gabrielsson (2002) contend that emphasising notation without performing by ear may well restrict overall musicianship.

In recent years, a growing literature has reported on aural learning practices, often spotlighting ear playing, particularly in the spheres of popular music, traditional music, jazz and other world musics (for some seminal texts, refer to Bennett, 1980; Berliner, 1994; Finnegan, 1989; McCarthy, 1999; and Monson, 1996). More recently work has looked at how such learning practices can be adapted for formal music education. The main focus tends to have been on classroom contexts and related areas such as school teacher-education (see, for example, Abrahams et al., 2012; Allsup, 2004; Davis, 2005; Downey, 2009; Evelein, 2006; Feichas, 2010; Finney & Philpott, 2010; Finnegan, 1989; Folkestad, 2006; Gatien, 2009; Georgii-Hemming & Westvall, 2010; Gower, 2012; Green, 2001, 2008; Jaffurs, 2004; Karlsen, 2010; Lebler, 2007; Lines, 2009; Mans, 2009; Marsh, 1999; McPhail, in press, 2012, 2013b; O’Flynn, 2006; Robinson, 2012; Rodriguez, 2004; Siefried, 2006; Vakeva, 2006; Vitale, 2011; Westerlund, 2006; Woody & Lehmann, 2010; Wright & Kanellopoulos, 2010).
The book *How popular musicians learn* (Green, 2001) identified key differences between the informal learning practices of popular musicians, and the approaches of formal music education. The former were characterised by:

- music the learner selected by him- or herself;
- copying audio recordings by ear, without notation;
- self- and peer-directed learning, typically in the absence of adult guidance;
- holistically-acquired skills rather than those resulting from a curriculum progressing from simple to more complex material;
- highly-integrated activities of listening, playing, composing and improvising.

This categorisation fed into a major national project for music classrooms in the United Kingdom, Musical Futures, which was funded by the Paul Hamlyn Foundation (www.musicalfutures.org; D’Amore, 2011; Green, 2008; Hallam, Creech, Sandford, Rinta, & Shave, 2008; Price, 2005). Musical Futures originally took place as a research-and-development project in secondary school classrooms in England, and has now started up in the other UK countries and in Australia, Canada, Singapore and elsewhere (e.g. Chua, 2013a, 2013b; Chua & Ho, 2013a, 2013b; Costes-Onish, 2013; Gower, 2012; Ho, 2013a, 2013b; Jeanneret, McLennan, & Stevens-Ballenger, 2011; McPhail, 2013a; Narita, 2012; O’Neill & Bespflug, 2012; Wright, 2011, 2012). The project involves a range of activities including classroom work-shopping as well as informal learning. At the core of the informal learning strand is a five- to six-week module called ‘Into the Deep End’ in which students collaborate in friendship groups, decide upon their own music and instruments, and engage in ear playing from audio recordings. Teachers respond to students’ needs by, for instance, modelling, questioning, providing help with finding pitches, making suggestions for holding instruments and posture, technique and many other aspects. This work has fed directly into the focus of the current article.

Following the Musical Futures project, Green mounted a new UK research venture called the Ear Playing Project (EPP) funded by the Esmée Fairbairn Foundation (see http://earplaying.ioe.ac.uk). The EPP aimed to bring informal learning practices, focussing on playing by ear from a recording, into the sphere of specialist instrumental teaching as distinct from the classroom environment. The project was designed primarily for classically-trained
instrumental teachers working in one-to-one or small group settings; particularly those who felt insecure in the realm of ear playing. The main aims were:

- the introduction of ear playing of which students might be unaware;
- the enhancement of aural skills;
- the development of a skill upon which students might build at home, thus potentially raising learner autonomy and motivation; opening a doorway to self-selected music and informal learning;
- allowing students to approach a range of music more creatively;
- offering instrumental teachers the opportunity to encounter and reflect on new ways of teaching and learning.

Data were collected quantitatively and qualitatively through questionnaire, interview and other means, as will be explained below. Although in this article we do make reference to some of the above aims, and to the participant teachers’ and students’ perspectives from our questionnaire and interview data, the central spotlight here is a ‘case-control’ experiment that was used to assess aural development outcomes from the EPP pedagogical strategies. Other elements of the project are discussed in Baker (2013), which reports on qualitative and quantitative findings; Green (2012b), and Varvarigou and Green (in preparation), which examines the concepts of ‘learning styles’ and ‘learning strategies’ in relation to the students’ responses; Varvarigou (in preparation) and Green (2012a), which focus more on teaching strategies and the role of the teacher; and Green (2014), which is a handbook for instrumental teachers, classroom teachers and band leaders.

For the purposes of the present article, our main research questions were:

- To what extent can we find any measurable improvements in students’ aural abilities following the use of the EPP pedagogical strategies, through a controlled, experimental design and independently-assessed outcomes?
- To what extent are teachers’ and students’ expressed opinions on the aural development resulting from the strategies mirrored by the results from the experiment?
We also consider some of the possible implications for instrumental teaching, given that lessons traditionally focus primarily on notation.

**The EPP Strategies**
A one-year pilot study took place (2009–10) involving 15 students, 4 teachers and observations of 104 lessons (see Green, 2012a, b). For the main study, following recruitment stands at various conferences in England, plus advertising on music education and other organisations’ web pages, 144 teachers attended induction sessions. These took place between September 2011 and April 2012. They included events for: Croydon Music and Arts; the Royal Northern College of Music (RNCM); the Royal National Institute of Blind People (RNIB); Forest Hill School (a Musical Futures Champion School); and six events held at the Institute of Education, University of London. At the training sessions, findings from the pilot study were shared, the teachers engaged in the ear-playing strategies themselves, and they were introduced into how to use them if they wished to join the research phase of the project. For the 135 teachers who elected to join, the strategies were to be implemented with a selection of their students, for a minimum of 10 minutes per lesson for a period of seven to 10 weeks. Teachers were also asked to be involved in a range of questionnaires, interviews, observations, audio-recordings, and also to allow their students to take part in these.

The teaching-and-learning approach was divided into three parts. During Stage 1, the focus was on ear-playing itself rather than informal learning. Students listened to a specially-composed pop-funk track which was provided on a project CD or downloaded from our website. They were then asked to play along with the bass riff heard on its own, by ear. They were required to find the notes on their instrument without notation. In keeping with Musical Futures (see above), the initial stage involved extensive independent learning, with the educator providing input only when he or she deemed it absolutely necessary. After the student had enough time to work out the notes for themselves, according to the teacher’s judgment, a range of teaching strategies were then implemented (see Green, 2012a, 2014). Other audio tracks separately provided the various component lines of the piece; thus, the student could gradually absorb it by ear. Stage 2 normally began at the third or fourth lesson. Students were asked to select one of six classical pieces, or excerpts thereof, by Bach, Handel, Mozart, Beethoven, Clara Schumann and Brahms. In a similar fashion to Stage 1, the students embarked on learning the piece by addressing its various component lines in response to our project CD. At Stage 3, around the fifth or sixth lesson, aspects more readily
associated with the holistic and autonomous elements of informal learning, as distinct from just aural learning, were utilised. Students could choose a piece of their own in any style, usually brought to lessons on a CD or iPod. They and the teacher would work out a portion of the piece by ear, this time, of course, without the help of the parts being made into individual tracks.

An Overview of Findings from the Questionnaires and Interviews
Overall, 135 teachers completed consent forms for continued involvement beyond the EPP induction sessions. Information subsequently sent to us by 54 of these participants indicated that at least 340 students, mostly taught individually or in pairs, worked with the project’s strategies. In addition, teachers took the strategies to 17 group-learning settings (bands, guitar groups and such like); the students in these groups were experiencing classical training. Quantitative and qualitative feedback was gathered through 13 teacher interviews and 54 teacher questionnaires. There was also a blog linked to the project website, and with the respondents’ permission online comments as well as e-mails also became part of our NVivo9 software database. Additionally, there were 42 student interviews and 193 student questionnaires. The quantitative items on our questionnaires included five-point rating scales and open questions to elicit reflective written comments and explanations; numerical data were stored using SPSS software. Findings and the methodology used are reported in full in Baker (2013). Our analysis revealed that the teachers who returned data considered the ear playing to be largely beneficial in a range of ways, although a sizeable number felt it had no significant impact in areas of pulse and rhythm (see Table 1).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Total questionnaires</th>
<th>Percentage in agreement</th>
<th>Number of responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My pupils developed a better ability to identify and find pitches on their instrument</td>
<td>87</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>The general aural development of my pupils improved</td>
<td>79.6</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>My pupils became more able to retain music in their heads (musical memory)</td>
<td>57.4</td>
<td>31</td>
<td></td>
</tr>
</tbody>
</table>
My pupils developed a better sense of pulse | 46.3 | 25  
My pupils developed a better sense of rhythm | 40.7 | 22

Similarly, 155 pupils thought it important to play by ear, as well as from notation (80.3% of the 193 questionnaires). In response to the statement ‘Would you prefer to be a professional musician who plays from notation, by ear, or both?’ 154 students (79.8%) remarked ‘Both’. A further 16 (or 8.3%) wished to be a musician who played only by ear. Our coding and analysis of the qualitative material was in accord with this positive standpoint on ear playing. For example, students made comments such as:

Well, I think one of the reasons I got Distinction is because it really helped with my aural. …I really thought it helped my aural tests in the exam and I’m really glad I did it. (Piano student, interview, 11 years old)

I can see it will help my aural like when I do my exams so… I think it’s quite important because it helps your aural when you do grades… (Piano student, interview, 13 years old)

You want to know all the answers to the [aural] exam… It’s very important in that case. (Clarinet student, interview, 10 years old)

The teachers agreed:

I think that they have all developed their listening skills, totally. (Woodwind teacher, interview)

[I have learned] just how positive it is to be free from notation…and how freeing [instrumental students] from the notation by getting them to play by ear is such a very obvious way of developing their aural skills. (Piano teacher, interview)

There were even remarks that indicated gains in musical memory:
Well, maybe [I now have] a longer memory, a playing memory. So, instead of remembering eight notes I remembered like 10 or 11 kind of thing. (Piano student, interview, 14 years old)

The Focus of the Current Article: The ‘Case-Control’ Experiment

The questionnaires and interviews clearly offered a positive view of the EPP strategies in terms of aural development. However, the researchers wished to test these findings in a more objective way than by relying in teachers’ and students’ observations and views. This was an investigation in which the teachers had opted to participate. That, in turn, might have been indicative of prior beliefs that ear playing was a worthy venture in educational terms; these values might be transmitted from the teacher to his or her students too. There could have been a ‘volunteer bias’ and a Hawthorne-type effect (see Cohen, Manion & Morrison, 2000; Landsberger, 1958); the latter might have been due to the ‘specialness’ of joining a prominent, national project. We wanted to ascertain whether or not the opinions of the teachers and their students (from the questionnaires and interviews) harmonised with more objectively-assessed learning outcomes. In doing so, we wanted to create potential for between-methods triangulation. We felt that any divide between qualitative and quantitative methodologies is blurry and the “measurement of all useful quantities requires a prior consideration of theory leading to the identification of a quality to be measured” (Gorard & Taylor, 2004, p. 5). In this vein, Haslam and McGarty (2003) have written of ‘cumulativeness’, or that effective research draws from and attempts to develop existing knowledge. It was our aim here to: seek points of intersection between methods in terms of the findings; develop our explanations further through complementary methods; and to show ‘fissures’ or ‘dissonances’ (Perlesz & Lindsay, 2003) between the outcomes of the various methods we selected. The latter would, we believed, lead to further lines of enquiry.

Method

The sample.

Our ‘case-control’ experiment used 16 matched pairs of students taught by four teachers. The 32 students were selected and paired by the teachers, according to their instrument and their Associated Board of the Royal Schools of Music (ABRSM) grade levels.² There were: five flautists, three clarinettists, two alto saxophonists, and the rest were pianists. They were taught in a mixture of group, one-to-one, school and home contexts. The students ranged from 10 to 14 years and were of Grades 1 to 5. The distribution of instruments was broadly
consistent with the overall EPP wherein 63.4% \((n = 206)\) of the participating students learned the piano. In each pair, one student belonged to an ‘experimental group’, which underwent the ear playing strategies (the independent variable); the other was part of a ‘control group’ that did not. It is worth mentioning that ear playing was not part of the typical learning or home activities of any of the students. We asked the teachers themselves to select and pair the students for the investigation, for two main reasons. One was that we required students to be not only of a similar age, but as far as possible, a similar level of attainment and aptitude; and we felt teachers, who knew their students, were best placed to make judgements about those areas. Secondly, the teachers needed to liaise with parents who could decide whether or not their child would be in our ‘experimental’ group and undertake a course of ear playing. There was a danger, of course, that teachers selected pupils they thought better motivated or adept at playing by ear, which might have tainted the data. However, the teachers were asked not to do this and parental choice was deemed important for ethical reasons. Furthermore, the requirement to match age, attainment and aptitude as far as was possible, usually left teachers with little or no choice as to which pupils to select. This issue might be allayed in further studies in which the research team select candidates or, alternatively, they are randomised, which would require a large sample. Our aim was to compare the effect of the ear playing in terms of aural development according to specified criteria (the dependent variable, see below), and to do this against a ‘baseline’, that is, the initial performance of both groups.

**Circumstances of the testing.**

The ABRSM kindly provided the researchers with a booklet of musicianship tests. Although the tests were in keeping with ABRSM practical musicianship testing, they were not publically available. The researchers had requested whether it would be possible to use ABRSM materials as the Board is the largest examining body of its type worldwide; it has a track record of effective aural testing which is integral to its syllabus; and we also wanted to have a system of assessment that was meaningful for, and familiar to the instrumental teachers. The participating teachers had all entered students for the Board’s examinations in the past. Fiske (1992) has argued that experimental research in music education sometimes leads to “complaints from non-researchers about the relevance of certain research areas to the day-to-day concerns of music teachers; there is some concern that the ‘real’ problems of music teaching are not considered by researchers” (p. 57). In each test, the pupil was played a recording of a melody, repeated twice over, and preceded by the key-chord. They were then told what key the music was in, and what the name of the first note was. The task was then to
play back the melody on their instrument. No notation was given. Although in normal exam circumstances, the ABRSM examiners play the melody live, the research team opted to deliver the tests via an audio recording so as to ensure exact consistency when it came to panel-assessment later on, and also because two research officers would undertake the testing and might therefore play the melodies slightly differently to each other. For Grades 1–3, the excerpts were two bars long and, for Grades 4–5, they were four bars. Each test was preceded by the standard ABRSM rubric:

I’ll play a short melody twice. Then I want you to repeat it from memory on your instrument. Here is the key chord [*named and played*] and your starting note [*named and played*].

Where the student played an instrument that was not of concert pitch (e.g. B-flat clarinet), the transposed key-chord and starting-note names were stated. The testing took place both in schools and the instrumental teachers’ homes between November 2011 and March 2012. The teachers volunteered to participate in the testing exercise as part of the wider EPP research. They were aware that the results would be reported with utmost confidentiality; the ethical procedures and arrangements for informed consent were followed according to the Guidelines of the British Educational Research Association (BERA).

The students’ responses to the tests were audio recorded. Students in each group (control and experimental) were given a pre- and post-test; the latter, the post-test, occurred after a period of seven to 10 weeks. During this time the students in the control group undertook the EPP strategies outlined above; the students in the experimental group carried on with their normal instrumental lesson.

**The appraisers.**

Fifteen teachers who had taken part in the EPP project formed the bulk of the expert panel of appraisers. They were all experienced instrumental teachers; some taught privately in their homes and others worked in primary or secondary schools. Nine of them assessed one half of the tests at one meeting and the other six assessed the other half at a separate gathering. This was not ideal; we would have preferred all the appraisers to assess all the tests; however in practice, owing to teachers’ availability and time constraints, splitting the exercise in this way proved necessary. In addition to the authors, Maria Varvarigou who was a research officer on
the project; Nigel Scaife, the ABRSM Syllabus Director; and John Holmes, the ABRSM Head Examiner also completed the exercise, and they assessed all of the tests. The researchers completed the exercise at the same meetings as the teachers. In all cases, the appraisers listened two or three times to each audio track of each student’s performance, and completed an appraisal document. They had no knowledge of whether or not the audio file was a pre- or post-test, or whether it was played by a control-group or an experimental-group student. Furthermore, they were given no information about the teacher, student or lesson context.

**The assessment criteria and the testing sheet.**

In preparation for assessment, the collected MP3 files were trimmed such that the rubric, starting note and key chord were removed using Audacity software. This was so that appraisers did not have to listen to these parts before each student’s response. An appraisal sheet was created. This included five, five-point rating scales, with 1 = ‘Very poor’, 2 = ‘Poor’, 3 = ‘Medium’, 4 = ‘Good’ and 5 = ‘Very good’, for the following criteria:

- pitch accuracy;
- contour accuracy;
- rhythmic accuracy;
- closure;
- tempo accuracy.

Whereas most of these terms are already largely understood between different educators and musicians, in some areas we gave a little explanation. The appraisers were told that ‘Closure’ referred to whether or not, for example, the candidate’s playing became erratic half way through and then came to a standstill; or, having stopped, whether or not he or she recovered and attempted, for example, the final two notes. ‘Tempo accuracy’ was distinct from ‘Rhythmic accuracy’ in the sense that the student might faithfully reproduce the rhythm, but at a slower pace than the recorded excerpt. All the assessment criteria were explained on a cover sheet that accompanied the assessment document; these were also discussed with the appraisers during the end-of-project meetings and any questions answered. Additionally, we asked the appraisers to give an ‘Overall grade’ (Grade A–E) in keeping with the other
criteria. These grades were later given numerical values for our analysis (with E = 1, D = 2, C = 3, B = 4 and A = 5).

ABRSM level descriptors were also included on the document (see Figure 1). These were designed in consultation with the Board and, as such, were a highly familiar, relevant format for the instrumental teachers concerned. It should be noted that there was extensive discussion between the ABRSM Syllabus Director and the researchers concerning the formation of all of the criteria used; although not intended as an exhaustive list of aural-skill elements, they were agreed as pertinent to the topic of investigation by the appraisers. Scrutiny of assessment criteria has been recommended as a principle of good statistical research (Haslam & McGarty, 2003).

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>17–18</td>
<td>Very good – Distinction</td>
<td>Fluent, accurate, and a prompt response</td>
</tr>
<tr>
<td>15–16</td>
<td>Good – Merit</td>
<td>At tempo, with some inaccuracy, omission or hesitation. Musical details conveyed</td>
</tr>
<tr>
<td>12–14</td>
<td>Medium – Pass</td>
<td>Outline in place, despite inaccuracy and hesitation</td>
</tr>
<tr>
<td>9–11</td>
<td>Poor - Fail</td>
<td>Uncertain and an inaccurate response. Frequent errors</td>
</tr>
<tr>
<td>6–8</td>
<td>Very poor – Fail</td>
<td>Very hesitant and a mostly incorrect or incomplete response</td>
</tr>
<tr>
<td>0</td>
<td>No work offered</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. ABRSM level descriptors

**Statistics.**

The authors conducted a split-plot repeated-measures analysis of variance procedure (F, ANOVA) with the raw data for each of the rating scale items. It was a ‘repeated measures’ procedure as it looked at two points in time, pre-test scores and, then, their post-test counterparts. The test could be considered a ‘split-plot’ because it compared the change in the mean and score distributions (from the pre- to post-test state) for both groups; the idea was to ascertain whether or not the difference between groups in that regard, the ‘interaction’, divergent enough such that it did not occur by chance alone. In tandem with this, we ran
Box’s $M$, Levene’s test and Mauchly’s $W$ with a Greenhouse-Geisser correction (denoted by $\bar{\varepsilon}$) as necessary post hoc tests.

The researchers believed that, even if this did not yield statistically-significant results, the shift in mean scores for the two groups would be revealed. This would offer compelling information and be helpful as a platform for future research in this vein; it would guide the formulation of a sturdier research instrument. Reflections in this regard are provided at the end of this paper.

**Findings**

Table 2 summarises the aggregated data. Here the standard deviation ($SD$) is the average distance of the raw scores from the arithmetic mean ($\bar{x}$), which represents the average score.

**Table 2. Results from the rating scales**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Group</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Shift in Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean, $\bar{x}$ &amp; SD</td>
<td>Mean, $\bar{x}$ &amp; SD</td>
<td></td>
</tr>
<tr>
<td>Pitch accuracy</td>
<td>Control</td>
<td>2.52, 1.02</td>
<td>2.53, 1.08</td>
<td>+ 0.01</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>2.74, 1.00</td>
<td>2.84, 1.00</td>
<td>+ 0.10</td>
</tr>
<tr>
<td>Contour accuracy</td>
<td>Control</td>
<td>2.89, 1.10</td>
<td>2.76, 1.19</td>
<td>- 0.13</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>2.88, 1.13</td>
<td>3.07, 1.00</td>
<td>+ 0.19</td>
</tr>
<tr>
<td>Rhythmic accuracy</td>
<td>Control</td>
<td>2.67, 1.01</td>
<td>2.76, 1.16</td>
<td>+ 0.09</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>2.66, 1.08</td>
<td>3.17, 1.20</td>
<td>+ 0.51</td>
</tr>
<tr>
<td>Closure</td>
<td>Control</td>
<td>2.63, 1.25</td>
<td>2.51, 1.18</td>
<td>- 0.12</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>2.59, 1.29</td>
<td>3.00, 1.27</td>
<td>+ 0.41</td>
</tr>
<tr>
<td>Tempo accuracy</td>
<td>Control</td>
<td>3.06, 1.07</td>
<td>3.10, 0.97</td>
<td>+ 0.04</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>3.15, 1.05</td>
<td>3.39, 0.98</td>
<td>+ 0.24</td>
</tr>
<tr>
<td>Overall grade</td>
<td>Control</td>
<td>2.65, 1.01</td>
<td>2.74, 1.11</td>
<td>+ 0.09</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>2.72, 1.06</td>
<td>3.06, 0.96</td>
<td>+ 0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Grade D)</td>
<td>(D)</td>
<td></td>
</tr>
<tr>
<td>ABRSM level</td>
<td>Control</td>
<td>11.57, 2.85</td>
<td>11.07, 3.14</td>
<td>- 0.50</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>11.51, 2.90</td>
<td>12.41, 2.87</td>
<td>+ 0.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(D)</td>
<td>(C)</td>
<td></td>
</tr>
</tbody>
</table>
Figures 2–8 illustrate how, according to every rating-scale assessment criterion, the experimental group students (the ones who had followed the EPP strategies) were assessed as achieving greater gains than the control group students. In each case, the vertical axis represents the mean score ($\bar{X}$), with the dotted and unbroken lines being the control and experimental groups (pre- to post-test) respectively. The standard deviation ($SD$) is also given at the start and end of each line to indicate the level of score dispersion.

**Figure 2.** Pitch accuracy ($F = NS$)
Figure 3. Contour accuracy ($F = \text{NS}$)

Whilst ‘Pitch’ and ‘Contour’ (Figs. 2–3) may seem compelling graphically, and could easily whet researchers’ appetite for further enquiry, the results for the ANOVA procedure used here did not produce statistically-significant results ($F = \text{NS}$ or non-significant).

Figure 4. Rhythmic accuracy.
Split-plot results for the ‘Rhythmic accuracy’ criterion were statistically-significant, though, with $F(1, 291) = 6.142, \text{MSE} = 1.078, p < .05$ (sig. .014). Box’s $M$ had a significance level of .517; furthermore, Mauchly’s $W$ (with a Greenhouse-Geisser correction, $\bar{\epsilon} = 1.00$) and Levene’s test ($p > .05$) were suitable to claim significance (i.e. the assumptions of sphericity and equality of variances had not been violated). The ear playing strategies had necessitated playing along to audio-recorded material which neither stopped nor accommodated students by adjusting tempi, as would for example a live piano accompanist. Accordingly, it might have been that superior attention to rhythmic accuracy had been developed in order to keep up with the audio, thus resulting in gains for the ear players.

![Figure 5. Closure](image)

‘Closure’ also yielded significant results with $F(1, 291) = 7.329, \text{MSE} = 1.401, p < .05$ (sig. .007); $p = .342$ for Box’s $M$, $p > .05$ for Levene’s test, and Mauchly’s $W$ was favourable ($\bar{\epsilon} = 1.00$). Perhaps, through the ear playing, students had internalized a greater sense of tonal centre, improved fluency and gained the capacity to apply a coherent, melodic ending to their test responses. Teachers also reported gains in student confidence through the ear playing in the questionnaires and interviews; this might have also contributed to the students’ greater persistence when forming their responses. Indeed, in general, data for accuracy in ‘Pitch’, ‘Contour’, ‘Rhythm’ and ‘Closure’ (Figs. 2–5) do appear to harmonise with the questionnaire
data mentioned earlier (Table 1); although only the ‘interaction’ of the control- and experimental-group for ‘Rhythm’ and ‘Closure’ can be said to hold any statistical power.

Figure 6. Tempo accuracy ($F = \text{NS}$)

‘Tempo accuracy’ (Fig. 6) showed a greater increase in the ear players’ mean score, pre- to post-test, but, again, $F = \text{NS}$. This was also the case with the ‘Overall grade’ criterion applied by the assessors.
However, the ABRSM levels also exhibited significant results, with $F(1, 281) = 9.713$, MSE $= 7.057$, $p < .05$ (sig. .002). Box’s $M$ returned a significance of .313 and there were satisfactory results for Levene’s test ($p > .05$) and Mauchly’s $W (\bar{E} = 1.00)$. According to the mean score, too, the effect of the EPP approach was to take pupils from a ‘Fail’ to a ‘Pass’
classification. Since the application of assessment rubrics such as these was most familiar to the appraiser pool, this was a captivating result. It is also worthy of note that these results occurred after students in the experimental group had been using the EPP strategies for only about 10 minutes, once per lesson, across approximately seven weeks. However, readers should be aware that this was merely a gain of 0.9 within a system that ranged from 0 to 18 marks. Also, students in the control group got worse by the greatest amount of all the criteria investigated, which was unexpected, and for which we can offer no explanation here.

Summary and Reflections
In summary, key points are:

1. In terms of the mean score for every criterion assessed, the ear players surpassed those who continued only with notation over the seven to 10 weeks of tuition between the pre- and post-test.
2. Regardless, the results are mixed with regard to statistical significance (and the $F$ statistic), with only accuracy in rhythm and melodic closure plus the ABRSM level descriptors showing a dramatic enough ‘interaction’ between the control and experimental groups.
3. Gains in the ear players for this experiment, at least in terms of the means, do seem to be largely supported by quantitative and qualitative data from EPP questionnaires (see Table 1). Nonetheless, this is clearly not the case for appraisals of rhythm, which had mixed results from the questionnaires (with 40.7% reporting improvements in students as compared to this dynamic in our case-control study, see Fig. 4).

Playing by ear, even over a short period of time, was clearly beneficial for ‘Closure’, which relates to students’ capacity to retain and reproduce a tonal centre, even if the thematic material is not entirely accurate in some cases. Our experiment also showed that it is important for ‘Rhythmic accuracy’ and these are both notable ingredients for fluency when responding to a sight-reading task. Our findings do appear to harmonise with connections held between ear-playing and sight-reading elsewhere, therefore (e.g. McPherson, Bailey, & Sinclair, 1997; also Luce, 1958; Sperti, 1970; McPherson, 1993/1994). Sight-reading proficiency in itself, even managing short fragments of material in the instrumental lesson, empowers students to absorb notated material and progress quickly, and this is surely
motivating for them. Students may become more able to copy their teachers’ demonstrations in lessons as a result of ear playing. Thus this critical listening can occur before relating it to their music notation. We agree with McPherson’s (2005) remark that skill acquisition is best served by a range of learning strategies: In this sense, ear playing is meaningful as part of a matrix of related activities for instrumental learners (including ear playing, sight reading, improvising, composing and performing from notation). An absence of enough deep, critical listening (e.g. to teachers’ renditions, live performances, audio recordings) and imitation of these could be one reason why some pupils fail to progress, as a result of inadequate aural development, and then ‘switch off’; the EPP created a situation which was designed to redress this a little. When we consider the relationships amongst ear playing, various aspects of aural development (as illustrated by our tests), sight reading, wider progress and motivation, and indeed, every type of musical performance, it is difficult to disagree with Priest’s (1988, 1989) idea that all musicianship is ‘by ear’. One of our questions concerned the possible implications of the strategies, given that instrumental lessons traditionally focus primarily on notation. The findings suggested palpable aural benefits in certain domains (i.e. ‘Rhythmic accuracy’, ‘Closure’ and the ABRSM level descriptors) as a result of injecting ear playing into lessons. These results, coupled with our EPP respondents’ views (our ‘data triangulation’), plus the literature cited in this article, may make a persuasive case for its inclusion in weekly lessons.

**Reflections on further research.**
A longer period is needed from the pre- to post-test, with a larger sample of students, in order to establish stronger connections between ear playing and other aural and instrumental skills. The implementation of the pedagogical strategies for only seven to ten weeks was the result of resource limitations. There may have been a number of factors of which we were unaware, and which could come to light over a longer period. Replication of these findings through additional research would throw light on their trustworthiness. The experiment reported in this article falls within the larger EPP and is limited to measuring musical skills rather than those associated with independence and autonomy (reported separately in Baker, 2013); further work could explore any connections between playing by ear, autonomy and motivation.

There was a response bias towards pianists in the EPP and this was also the situation with our case-control experiment. It would be interesting to run these tests again with children of other
instrumental disciplines. There may be idiosyncratic responses to them, too, due to the technicalities of specific musical instruments, or even the pedagogical approaches taken by teachers of those instruments. Owing to the standards of the pupils chosen for the testing, upper ABRSM grades are absent from the sample. Our teachers selected candidates for the testing and this is indicative of their personal interest in developing students’ ear playing and aural capacities from an early point. Whilst limiting grade standards focussed the sample, it reduced the generality of findings in terms of making practical recommendations for instrumental teaching at large. Haslam and McGarty (2003) suggest that ‘trade-offs’ and ‘uncertainties’ are unavoidable in statistical methods: “if we reduce descriptive uncertainty by using a very homogeneous sample…we may increase external uncertainty about the applicability of our findings to a broader population” (pp. 425–426). The way in which candidates prepare for testing is an important consideration in such designs too (see Cohen et al., 2000). It would be interesting to see if our present results are replicated by alternative methods under different circumstances. As with much research, the results provided compelling insights but, in tandem, opened significant windows for future enquiry.

**Funding**

The project was funded by the Esmée Fairbairn Foundation, UK in 2009–10 and 2011–12 under their ‘New approaches to learning’ strand [grant number 09-3272].

**Notes**

1. Organisations who kindly helped the research team by promoting the research were the Associated Board of the Royal Schools of Music (ABRSM), the European Piano Teachers Association (EPTA), the European String Teachers Association (ESTA), and the Incorporated Society of Music (ISM) to name a few. Advertisements were placed on websites and in newsletters. The research officers also attended ABRSM conferences in London, Manchester and Birmingham.

2. For further information on ABRSM grade levels and how they relate to QCF (Office for Qualifications and Examinations Regulations) national standards in the United Kingdom, see [www.abrsm.org](http://www.abrsm.org)

**References**


during one-to-one instrumental lessons.


Author Biographies

David Baker was a music teacher in England for many years. He currently works as a Research Associate at the Institute of Education, University of London and as a Course Tutor in the MA programme in instrumental teaching at Reading University, UK. David has published research on teachers using biographical methods and, more recently, chapters in Developing the musician (Stakelum, 2013) on ear playing and conceptions of musical ability. He is Principal Investigator for the ‘Visually-impaired musicians’ lives’ project at London’s Institute of Education, which is funded by the Arts and Humanities Research Council, and supported by the Royal Academy of Music and Royal National Institute of Blind People.

Lucy Green is Professor of Music Education at the London University Institute of Education, UK. Her research interests are in the sociology of music education, specializing in issues of meaning, ideology, gender, popular music, informal learning, and new pedagogies. Lucy created the research and development project “Informal Learning in the Music Classroom” within the British venture “Musical Futures”, and her current research is taking this work
forward into instrumental tuition. Her work has been widely published in seminal books and journal articles. She is the author of four books on music education, and the editor of *Learning, teaching and musical identity: Voices across cultures* (Indiana University Press, 2011). Her next book will be published in 2014 by Oxford University Press, entitled *Hear, listen, play! How to free your students’ aural, improvisation and performance skills.*