Educational benefits of using game consoles in a primary classroom: A randomised controlled trial

David J. Miller and Derek P. Robertson

David J. Miller is Senior Lecturer in Education at the University of Dundee. His research interests include primary pedagogy, and in particular the relationship with personal and social development. Derek P. Robertson works at Learning and Teaching Scotland, Dundee. He is National Adviser for Emerging Technologies and Learning, and has a particular interest in the potential of game-based learning in school. Address for correspondence: Dr. D. J. Miller, School of Education, Social Work and Community Education, University of Dundee, Nethergate, Dundee DD1 4HN, UK. Email: d.j.miller@dundee.ac.uk

Abstract
It is known that computer games are motivating for children, but there is limited direct evidence of their effects on classroom learning. The studies that are available tend to be limited in terms of output data reported, or small in scale, or both. The aim of this randomised controlled trial was to upscale a recent study by Miller and Robertson investigating the effects of a commercial off-the-shelf computer game on children’s mental computation skills and self-perceptions. A pre-post design was employed, with 634 primary (elementary) school children (10–11 years old) from 32 schools across Scotland. Schools were randomly assigned to experimental or control conditions. In the experimental schools, children used a games console for 20 minutes each day, running a ‘brain training’ game. The controls continued with their normal routine. The treatment period was 9 weeks. Significant pre-post gains in accuracy and speed of calculations were found in both experimental and control groups over the treatment period. Gains in the experimental group were 50% greater than those of the controls in accuracy, and twice those of the controls in speed. There were no significant changes in two measures of self-concept for either group. There was a small but statistically significant improvement in attitude towards school among the experimental group but not the controls. When scores were analysed by ability, different patterns were apparent. The design of the study allows a degree of confidence when generalising from these results. Some implications of the findings are discussed.

Introduction
Teachers and researchers have been interested in the use of game-based applications in schools for many years now (see, for example, Facer, 2003; McFarlane, Sparrowhawk & Heald, 2002; Randel, Morris, Wetzel & Whitehill, 1992). Many games are specifically written for education purposes, (educational software or e-learning games), but there has also been increased interest in the potential of commercial off-the-shelf computer games (COTS) in the classroom (Kirriemuir & McFarlane, 2004). The arguments for computer games rest on beliefs about enhanced knowledge and skills, and associated improvements in attitudinal factors and engagement (Passey et al, 2004; Sandford, Ulicsak, Facer & Rudd, 2006).

Theoretical explanations for the popularity of gameplaying can be found in both psychological and sociocultural theories. Among the former, the notion of ‘flow’ (Csikszentmihalyi, 1975,
1990; Malone, 1980) is prominent. The concept of flow refers to a feeling of complete absorption or engagement in an activity, of being ‘in the zone’. Analysing data gathered from interviews and surveys, Csíkszentmihályi identified several characteristics of flow, including a complete focus on key goals or targets, high levels of concentration on the task and a lack of self-consciousness: ‘losing oneself’ in the activity. Central to this is a match between ability and challenge, and the receipt of immediate feedback, allowing rapid adaptation to the situation and modification of response. These factors create a sense of personal control over an activity, and often contribute to an individual experiencing a distorted sense of time. (See also Inal & Cagiltay, 2007).

In contrast, several authors (eg, Gee, 2003; Prensky, 2001; Sandford et al, 2006) have highlighted social and cultural factors in relation to playing computer games. Gee (2003) discusses semiotic domains, defined as ‘any set of practices that recruits one or more modalities [...] to communicate distinctive types of meaning’ (p. 18). Enthusiastic gamers will share a semiotic domain; that is, they will share norms, values and beliefs about what counts as worthwhile knowledge, what is good and what is not in terms of performance, and judgements about status within the domain of digital games. In particular, they have views about the nature of learning based on their experiences of such games. This particular domain may be contrasted with another semiotic domain—that which encompasses school work and classroom values (a domain possibly shared with teachers). The games domain reflects a set of values and expectations about what is good, engaging and worthwhile in the games, and it is important to realise that it is a semiotic domain that teachers may not share—or fully understand.

The investigation of gameplaying in school is fraught with many difficulties (Kirriemuir & McFarlane, 2004), and the evidence is neither extensive nor robust (Condie & Munro, 2007). Although much has been written about the benefits of games and games-based learning—particularly in terms of beliefs and attitudes of teachers, pupils and parents (eg, McFarlane et al, 2002; Sandford et al, 2006)—there are fewer studies reporting output measures in terms of attainment.

Reports and reviews on games-based learning
An early review of the evidence on gameplaying was conducted by Randel et al (1992). The authors examined 67 studies conducted prior to 1991, and reported mixed results, with more than half showing no link between gaming and attainment. Methodological weaknesses were identified in many studies. The area where the strongest links were seen was mathematics, but even here the authors commented on methodological problems in the majority of studies. A shift of focus is evident in the current educational literature, where much that has been written about the benefits of games and games-based learning—particularly in terms of beliefs and attitudes of teachers, pupils and parents (eg, McFarlane et al, 2002; Sandford et al, 2006)—there are fewer studies reporting output measures in terms of attainment.

One recent meta-analysis of computer gaming and interactive simulations for learning was conducted by Vogel et al (2006). While commenting that the research base was insufficient to draw conclusions with much confidence (p. 238), they reported that interventions using interactive simulations or games (or both) tended to result in higher cognitive gains and better attitudes towards learning compared with those using traditional teaching methods. It should be noted that the studies analysed focused on adults as well as children; it is not clear how many of them referred to children in school. As with the earlier review by Randel et al (1992) there was criticism of the quality of the studies. Although 248 studies were evaluated for inclusion, only 32 of these met the requirements and were included in the analysis: ‘methodological and reporting flaws are rampant in the unused articles’ (ibid.). They specifically criticised the absence of control groups, lack of statistical data, the omission of important demographic details, and interventions which were not described in sufficient detail.

All of these factors suggest the need for caution in interpreting the research literature in the area—a view echoed in a landscape review of the impact of information and communication technology in schools (Condie & Munro, 2007). The authors concluded that research evidence in relation to games and learning was very limited—for both e-learning and COTS games. The evidence that does exist points to benefits of gameplaying in terms of faster processing of information, enhanced selection of relevant material and high levels of engagement and interest.

Recent work in gaming which reports performance data
Since the date of these reviews, some relevant studies have been conducted. Focusing on gameplaying in mathematics, Ke (2008) investigated the use of educational drill-and-practice computer games with 15 fourth and fifth grade (age 10–13) pupils over 5 weeks. Although there was evidence of more positive attitudes towards mathematics, there was no significant gain in mathematics performance. Apart from the very small sample size, it is important to note that there were no controls in this study.

A recent study which investigated gains from an educational computer game was conducted by Tüzün, Yılmaz-Soylu, Karakus, Inal and Kızılkaya (2009). The participants were 24 fourth and fifth grade pupils. Pre-post achievement test data indicated significant learning gains in geographical knowledge together with improved intrinsic motivation. Once more, a degree of caution is necessary when interpreting the findings. There were no controls and the participants had been selected by the school authorities, being above average in ability, levels of motivation and parental background. Additionally, three teachers worked with the group during the treatment period, and the quantitative data reported came from only 13 of the children.

Working with 71 primary school children (10–11 years old) from three classes, Miller and Robertson (2009) investigated the effects on mathematical computation and self-perceptions when children used a COTS programme on a game console over a 10-week treatment period. Significant pre-post gains were found in the game console group for both accuracy and speed of calculations, while results for two comparison groups were mixed. One comparison group showed significant gains in speed but not accuracy; the other showed gains in accuracy but not speed. In each case, the gains were less than half those achieved by the game consoles group. In addition, the game consoles group showed significant gains in global self-esteem, but not in other aspects of self-concept. No gains in self-perceptions were evident in the two comparison groups. Although this study demonstrates ecological validity, the authors acknowledged incomplete information about the experiences of one of the comparison groups, and advised caution when interpreting their scores. No such concerns were expressed about the scores of the other comparison group, effectively a no-treatment control group. Miller and Robertson commented on the dangers of generalising from the findings, given the small sample size.

Summary and rationale for current study
While encouraging messages emerge from the literature on games-based learning, the evidence in terms of performance data is limited. The need for a cautious interpretation of the evidence is reinforced by reviews which have pointed out methodological shortcomings in studies. The current investigation was an attempt to overcome some of the weaknesses in previous work. It was essentially an upscaling of the work of Miller and Robertson (2009). It attempted to address the limitations of that study—its small scale and the nature of the comparison groups—in order that findings could be generalised more widely. Specifically the aims were to investigate the effects of the game console on accuracy and speed of computation, pupil self-perceptions and attitudes towards school.
Method
Overview
A randomised controlled trial was conducted, employing pre-post measures of mathematics performance and attitudinal factors. The aim was to investigate changes in children’s mental computation performance and self-perceptions over a 9-week period. In the experimental classes, the children used a game console every day for 20 minutes: a Nintendo DS Lite (Nintendo Co. Ltd., Kyoto, Japan), playing Dr. Kawashima’s Brain Training. The controls continued with their regular class programme.

The main questions being investigated were:
1. What changes are evident in children’s accuracy and speed of computation?
2. What changes are evident in relation to self-concept and attitude to school?
3. To what extent are gender differences evident in the pattern of scores?
4. Does the game console have a differential effect on children who vary in terms of mathematics ability?
5. Are different patterns of scores evident in children who have experience of playing the game at home?

Participants
A stratified random sample was created in a step-wise manner. The first step was to identify schools in the lowest quartile in terms of socio-economic status (SES) in four education authorities in Scotland. This was achieved by consulting government statistics on free school meal entitlement, an accepted indicator of SES. From this pool of potential participants, 32 schools were randomly selected. Next, these schools were randomly assigned to experimental or control conditions. Each school in the experimental group was given a set of Nintendo DS lite game consoles for a primary 6 class.

The participants comprised 634 primary school (elementary school) children, 10–11 years of age. The 32 schools included small rural schools, mid-size town schools varying in terms of age and facilities, and larger inner-city schools. All were mainstream state schools, and children were the regular classes of the teachers involved. In practice, this meant that there was a range of abilities, from more able children, through those working within the expected levels on the official curriculum guidelines, to those receiving additional support. No individuals in the sample were excluded from the project on the basis of ability or other personal characteristics.

Instruments
In order to measure children’s speed and accuracy of mental computation, the 100-item ‘Number Challenge’ employed by Miller and Robertson (2009) was used. The authors had devised this instrument to be age- and stage-appropriate, focusing on calculations within Level D in the Scottish mathematics curriculum (Scottish Office Education Department (SOED), 1991). It included a range of addition, subtraction, multiplication and division items, from simple two-digit calculations to more advanced sums involving three two-digit numbers testing the associative property. Answers were marked simply as correct or incorrect, with the range of scores being from 0 to 100. Time taken to complete the challenge was also recorded. The maximum time allowed was 25 minutes, and this time was assigned to any child who had not finished at the end of the testing period.

It was also decided to investigate aspects of children’s self-perceptions. Miller and Robertson (2009) had reported gains in global self-esteem, but found no significant change in more specific measures of mathematics self-concept. Given what is known about the relationship between performance in a given domain and self-perceptions in that domain (eg, Marsh, 2008) this was a surprising finding. Simply stated, one might reasonably have expected to see improved perfor-
mance in mathematics reflected in some way in children’s mathematics self-concept. One possible explanation is that the instrument employed was not sensitive enough to pick up changes of that nature. Miller and Robertson had employed the Burnett Self Scale (Burden & Burnett, 1999) but in the current study it was decided to employ instead a shortened form of Marsh’s Self Description Questionnaire (Marsh, 1992).

This instrument measures self-perceptions in several domains. In its full form it comprises 76 items, organised into eight sub-scales. The instrument has good reliability for this age group; Marsh obtained \( \alpha \) coefficients ranging from 0.80 to 0.94 for the various scales with elementary school pupils. A shortened form was created by the current authors. This included items from two sub-scales only: mathematics and general-school sub-scales. These were interpreted as reflecting mathematics self-concept and general academic self-concept, respectively. While acknowledging that modifying a scale can affect the psychometric properties of the instrument, it did allow us to create a relatively short, user-friendly scale for the participants. (see Appendix A for amended scale.)

Following Marsh, mathematics self-concept is defined as the set of beliefs an individual holds about himself or herself as a mathematician; whereas, academic self-concept refers to an individual’s beliefs about abilities across a range of academic subjects. Two things follow from these definitions. First, it can be seen that the latter subsumes the former; this characteristic is reflected in a hierarchical model of self-concept (Marsh & Craven, 2006). Second, potentially at least, children’s beliefs about their performance in mathematics influence both mathematics self-concept and academic self-concept. The interrelation between these two measures, and their relationship with general self-concept is effectively summarised by Marsh and Craven (2006, p. 137).

Additionally, it was decided to investigate any change in children’s attitude towards school over the treatment period. A short scale was created based on items in a previously published instrument (Sa’di, 2001) designed to assess children’s attitudes towards primary school (see Appendix B). The author reported an alpha coefficient of 0.88 for the original scale, indicating high reliability. Items selected for the shortened scale were those which the researchers felt were culturally appropriate for the target group. An improved score on this scale would reflect a more positive attitude towards school.

In addition to these measures, extra information was collected from schools about children’s recorded level of ability in mathematics. These data were in the form of levels of attainment against national standards, previously recorded by the school. Information was also collected from children about their computer use at home, since it was felt that this might act as a moderator variable in children’s interaction with the COTS experience provided in the intervention.

**Procedure**

Prior to the treatment period, 1-hour training sessions were conducted for all participating teachers on a group basis. Although teachers expressed confidence after the training, they were also given contact details for the researchers so that ongoing advice and support were available if needed. In practice, the console has a low-skills threshold, and no teachers requested extra help. Pre-tests were conducted in April, and the experimental group began the intervention programme. Each week, experimental classes used *Dr. Kawashima’s Brain Training* programme for 20 minutes first thing each day, 5 days a week. On the fifth day, the participants also completed the ‘Brain Age Check’, a task which provided feedback about their performance. The treatment period was 9 weeks, with post-test data collected immediately thereafter.

The data collected were coded and entered into Statistical Package for the Social Sciences (SPSS, an IBM Company, Chicago, Illinios, USA), version 16. Cases were excluded list-wise; that is, any
children with missing data were excluded from the analyses. They were then subject to analysis of variance (ANOVA) techniques; more detail is provided in relation to each research question below.

**Findings**

A one-way between-groups ANOVA was conducted on the pre-test data to investigate the possibility of significant differences between the experimental and control groups at the start of the intervention.

There were no significant differences between experimental and control groups in relation to the dependent measures: number of sums correct \(F(1, 697) = 0.44, p = 0.51\); time taken \(F(1, 697) = 2.7, p = 0.10\); mathematics self-concept \(F(1, 697) = 0.004, p = 0.95\); academic self-concept \(F(1, 697) = 2.49, p = 0.115\); or attitude to school \(F(1, 697) = 1.35, p = 0.25\).

**What changes were evident in children’s accuracy and speed of computation?**

Mean pre-test and post-test scores were calculated for both groups, and mixed between-within subjects ANOVA applied to the data. One-tailed rather than two-tailed significance figures were calculated based on a priori assumptions about gains.

For accuracy of computation, there was a significant effect for time \(F(1, 633) = 71.9, p < 0.001\); both groups improved significantly over the intervention period. The effect size using eta-squared \(h^2\) was 0.10, a moderate effect. However, the interaction between time and condition was also significant \(F(1, 633) = 3.27, p = 0.035\); the rate of increase differed significantly for the two groups, with the experimental group improving at a faster rate. The effect size was \(h^2 = 0.01\), a small effect. This difference is reflected in Table 1 where the gain for the game consoles group is over 50% higher than that for controls.

For speed of processing, there was a significant effect for time \(F(1, 633) = 615.5, p < 0.001\), with \(h^2 = 0.49\), a large effect; both groups became significantly faster over the intervention period. However, the interaction between time and condition was also significant \(F(1, 633) = 83.34, p < 0.001\); the experimental group improved their time more than the controls over the treatment period. The effect size was \(h^2 = 0.12\), a moderate effect. It can be seen from Table 2 that the improvement for the experimental group was twice that of the controls (Table 3).

**What changes were evident in relation to self-concept and attitude to school?**

In relation to mathematics self-concept, there were no significant effects of time \(F(1, 631) = 0.08, p = 0.387\) or interaction with treatment condition \(F(1, 631) = 0.002, p = 0.484\). In each case \(h^2 < 0.01\). Children’s mathematics self-concept did not change significantly in either group over the treatment period. The data on academic self-concept show a similar pattern. There

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean (SD)</th>
<th>(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sums correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>358</td>
<td>77.17 (21.54)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>341</td>
<td>78.29 (22.76)</td>
<td></td>
</tr>
<tr>
<td>Time taken</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>358</td>
<td>18:44 (06:02)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>341</td>
<td>17:59 (05:52)</td>
<td></td>
</tr>
<tr>
<td>Academic self-concept</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>358</td>
<td>33.90 (7.15)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>341</td>
<td>33.05 (7.13)</td>
<td></td>
</tr>
<tr>
<td>Mathematics self-concept</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>358</td>
<td>36.13 (9.21)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>341</td>
<td>36.09 (8.57)</td>
<td></td>
</tr>
<tr>
<td>Attitude to school</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>358</td>
<td>17.58 (3.05)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>341</td>
<td>17.32 (2.97)</td>
<td></td>
</tr>
</tbody>
</table>

SD = standard deviation.
were no significant effects of time \( F(1, 631) = 0.86, p = 0.168 \) or interaction with treatment condition \( F(1, 631) = 2.08, p = 0.075 \). In each case \( \eta^2 < 0.01 \). Children’s academic self-concept did not change significantly in either group over the treatment period (Table 4).

In relation to children’s attitude to school, there was no significant effect of time \( F(1, 632) = 0.08, p = 0.398 \), but the interaction between time and condition was statistically significant \( F(1, 632) = 2.74, p = 0.049 \). In each case \( \eta^2 < 0.01 \). Children in the experimental group improved in their attitudes towards school; whereas, those in the control group did not. However, it should be noted that the effect size was very small.

To what extent were there gender differences in the pattern of scores?

Gender differences were investigated using a one-way between-groups ANOVA (two-tailed). Although no significant difference was found in terms of the number of sums correct at pre-test

\[
\begin{array}{lcccc}
\text{Condition} & N & \text{Mean (SD)} \\
\text{Intervention} & 326 & \text{Pre-test} & 78.56 (20.11) & \text{Post-test} & 83.04 (16.86) & \text{Gain} & 4.48 \\
\text{Control} & 309 & \text{Pre-test} & 78.74 (22.54) & \text{Post-test} & 81.65 (20.80) & \text{Gain} & 2.91 \\
\end{array}
\]

SD = standard deviation.

\[
\begin{array}{lcccc}
\text{Condition} & N & \text{Mean time (m : s) (SD)} \\
\text{Intervention} & 326 & \text{Pre-test} & 18:32 (06:06) & \text{Post-test} & 13:29 (05:45) & \text{Change} & -05.03 \\
\text{Control} & 309 & \text{Pre-test} & 17:54 (05:51) & \text{Post-test} & 15:34 (05:59) & \text{Change} & -02.20 \\
\end{array}
\]

SD = standard deviation.

\[
\begin{array}{lcccc}
\text{Condition} & N & \text{Mean} & \text{SD} \\
\text{Maths self-concept} & & & & \\
\text{Time 1} & \text{Intervention} & 36.46 & 9.037 & 327 \\
\text{Control} & & 36.13 & 8.612 & 306 \\
\text{Time 2} & \text{Intervention} & 36.53 & 9.042 & 327 \\
\text{Control} & & 36.21 & 8.969 & 306 \\
\text{Academic self-concept} & & & & \\
\text{Time 1} & \text{Intervention} & 34.06 & 7.113 & 327 \\
\text{Control} & & 33.31 & 7.145 & 306 \\
\text{Time 2} & \text{Intervention} & 33.95 & 7.344 & 327 \\
\text{Control} & & 33.83 & 7.471 & 306 \\
\text{Attitude to school} & & & & \\
\text{Time 1} & \text{Intervention} & 17.67 & 3.049 & 327 \\
\text{Control} & & 17.39 & 2.885 & 307 \\
\text{Time 2} & \text{Intervention} & 17.88 & 2.899 & 327 \\
\text{Control} & & 17.24 & 2.917 & 307 \\
\end{array}
\]

SD = standard deviation.

\[
\begin{array}{lcccc}
\text{Condition} & N & \text{Mean} & \text{SD} \\
\text{Intervention} & & & & \\
\text{Control} & & & & \\
\end{array}
\]
[F(1, 356) = 0.510, p = 0.476], there were significant gender differences in relation to all other measures: time taken [F(1, 356) = 13.002, p < 0.001]; mathematics self-concept [F(1, 356) = 6.311, p = 0.012]; academic self-concept [F(1, 356) = 7.767, p = 0.006]; and attitude to school [F(1, 356) = 12.520, p < 0.001].

Accordingly, analysis of covariance was employed to control for these pre-treatment differences. In each case, the independent variable was gender and the dependent variable was the score on the measure at the end of the intervention period. Participants’ pre-test scores on the same measure were used as covariates in each case.

After adjusting for pre-intervention scores, there were no significant differences between males and females on post-intervention scores: sums correct [F(1, 323) = 0.261, p = 0.610]; time taken [F(1, 323) = 0.001, p = 0.973]; mathematics self-concept [F(1, 324) = 0.189, p = 0.664]; academic self-concept [F(1, 324) = 1.516, p = 0.219]; and attitude to school [F(1, 324) = 3.078, p = 0.080]. In each of these cases η^2 < 0.01. Boys’ and girls’ scores did not differ on any measures as a result of their experiences (Table 5).

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Mean time 1</th>
<th>SD</th>
<th>Mean time 2</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sums correct</td>
<td>Male</td>
<td>142</td>
<td>79.84</td>
<td>19.881</td>
<td>83.62</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>184</td>
<td>77.57</td>
<td>20.275</td>
<td>82.60</td>
</tr>
<tr>
<td>Time taken</td>
<td>Male</td>
<td>142</td>
<td>0:17:06</td>
<td>0:06:38</td>
<td>0:12:27</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>184</td>
<td>0:19:38</td>
<td>0:05:24</td>
<td>0:14:18</td>
</tr>
<tr>
<td>Maths self-concept</td>
<td>Male</td>
<td>143</td>
<td>37.62</td>
<td>9.175</td>
<td>37.19</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>184</td>
<td>35.57</td>
<td>8.850</td>
<td>36.02</td>
</tr>
<tr>
<td>Academic self-concept</td>
<td>Male</td>
<td>143</td>
<td>32.92</td>
<td>7.489</td>
<td>32.73</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>184</td>
<td>34.95</td>
<td>6.949</td>
<td>34.90</td>
</tr>
<tr>
<td>Attitude to school</td>
<td>Male</td>
<td>143</td>
<td>17.02</td>
<td>3.251</td>
<td>17.28</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>184</td>
<td>18.17</td>
<td>2.789</td>
<td>18.35</td>
</tr>
</tbody>
</table>

SD = standard deviation.

Gain scores were calculated for both accuracy and speed of computation and were analysed according to children’s recorded level of attainment in national tests in mathematics. In these analyses, level A is the lowest level in terms of ability, and level E the highest.

Figure 1 indicates that in terms of accuracy, the less able children (level A) tended to improve more than the more able children. This was investigated using Pearson product–moment correlation. There was a small but significant negative relationship between the two variables (r = −0.29, n = 318, p < 0.001) the lower the ability level the greater the gain (Figure 2).

Regarding speed of processing, it can be seen that the nature of the relationship was different. It appears that the middle-ability children tended to improve their time more than the children at the top and bottom of the ability range. Again, this was investigated using Pearson product–moment correlation coefficient. As would be expected with such a pattern, there was no significant relationship between the two variables (r = −0.045, n = 318, p = 0.428).

Data collected from the children had indicated that many owned game consoles and COTS games; several stated that they had used Dr. Kawashima’s Brain Training at home. To investigate whether this additional experience had influenced how children performed, participants who claimed to have used the game at home were compared with those who had not. A between-groups ANOVA

indicated that the scores of these two groups did not differ on pre-test: number correct $[F(1, 312) = 1.143, p = 0.286]$; time taken $[F(1, 312) = 0.053, p = 0.818]$.

Mixed between-within subjects ANOVA indicated that there was a significant effect for time: accuracy $[F(1, 312) = 46.299, p < 0.001]$; and speed $[F(1, 312) = 471.04, p < 0.001]$. However, the interaction between time and condition was not significant: accuracy $[F(1, 312) = 1.112, p = 0.293]$; speed $[F(1, 312) = 0.543, p = 0.462]$. The performance of children who had extra
Experience of playing the game at home did not differ from those who lacked such experience (Table 6).

**Discussion**

The main findings from the study are that: (1) both groups improved in speed and accuracy of computation during the course of the intervention; and (2) the rate of increase differed for the two groups, with the experimental group improving at a significantly faster rate. These differences are reflected in higher gain scores for the game consoles group in each case. The findings are broadly comparable with the small-scale study by Miller and Robertson (2009). In that study, children following an intervention regime similar to the current study showed significant gains in both speed and accuracy; these gains were twice those of the nearest comparison group. However, the current study is based on a larger and more representative sample.

The picture in relation to attitudinal factors provides a contrast. There was no significant change in mathematics self-concept or academic self-concept in either group. It had been anticipated that increases in computational skill might have influenced children’s self-perceptions in these aspects of self-concept. Work in relation to the reciprocal effects model (e.g., Marsh & Craven, 2006) has highlighted the interrelationship of children’s performance in a given domain and their self-concept in that domain. In the current study, no such relationship was evident; the gains in performance in computation were not reflected in mathematical self-concept. Similarly, there was no significant change in academic self-concept, another aspect of self-perceptions where one might have expected to see a change. The fact that there were no such changes is consistent with the study by Miller and Robertson (2009). In that study, although gains were found in global self-esteem, in the more specific aspects of mathematics self-concept and learning self-concept, there were no significant changes.

In some respects, this remains a puzzling finding. Certainly, computation is only one aspect of mathematics, and an individual’s self-concept in this domain reflects a wider range of beliefs and judgements than simply performance on mental arithmetic tasks. For example, problem-solving ability may play a significant part in self-judgements here, whereas, the activities on the game consoles focused only on computation. On the other hand, computation is fundamental to success in mathematics in elementary school—a fact not lost on children—and one might reasonably expect improved performance to have some impact on self-beliefs in this domain. It may simply be that 9 weeks is insufficient time for self-perceptions to be changed. However, in informal discussion after the project ended, teachers invariably commented that children were more confident in—and often more enthusiastic about—mathematics work as a result of their experiences. Why therefore were these perceptions not reflected in self-concept scores?

<table>
<thead>
<tr>
<th>Experience of this game at home</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number correct time 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>77.30</td>
<td>21.565</td>
<td>183</td>
</tr>
<tr>
<td>Yes</td>
<td>79.78</td>
<td>18.366</td>
<td>131</td>
</tr>
<tr>
<td>No correct time 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>82.47</td>
<td>17.248</td>
<td>183</td>
</tr>
<tr>
<td>Yes</td>
<td>83.56</td>
<td>16.410</td>
<td>131</td>
</tr>
<tr>
<td>Time taken time 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0:18:35</td>
<td>0:06:13</td>
<td>183</td>
</tr>
<tr>
<td>Yes</td>
<td>0:18:45</td>
<td>0:06:01</td>
<td>131</td>
</tr>
<tr>
<td>Time taken time 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0:13:42</td>
<td>0:05:57</td>
<td>183</td>
</tr>
<tr>
<td>Yes</td>
<td>0:13:31</td>
<td>0:05:37</td>
<td>131</td>
</tr>
</tbody>
</table>

SD = standard deviation.

Table 6: Analysis by children’s experience of playing the game

Three possible explanations suggest themselves. The first is that in the informal discussions with teachers and children, researchers were receiving messages which did not accurately reflect genuine changes in beliefs. The second explanation is that the self-concept scales employed were not sensitive enough to capture this change in beliefs. The fact that the instrument was modified might lend credence to that explanation, although the two sub-scales most relevant to mathematical performance were used in their entirety, and the instrument is regarded as the benchmark measure in this area. The third explanation is one which relates to domains of experience, or semiotic domains (Gee, 2003). Adopting such a perspective, we might interpret events as follows. Children were aware of the fact that they were getting better at playing the game, and possibly that this involved improvements in number bonds and speed of responding. However, in the children’s minds, these skills were situated in—and identified with—the domain of computer games. This is distinct from the domain of ‘school work’, and despite the fact that the games were used in the classroom, there may have been little or no transfer to the school domain. Of course, this is speculation at this point in time, but it is certainly worthy of further investigation, because of the educational value of children fully appreciating what they have achieved.

There was a slight but statistically significant improvement in attitude towards school in the experimental group, but not in the controls. The effect size here was very small, and it is important not to overestimate the importance of this. However, the scale consisted of only five items and the period of the intervention was relatively short at 9 weeks. It would be interesting to investigate whether an extended scale would produce a clearer picture here, particularly if the treatment period was longer.

In terms of differential effects on children, several issues were of interest. First, there were no significant gender differences. Although Inal and Cagiltay (2007) reported gender differences in young children playing computer games, including the finding that boys experienced more flow experiences, the meta-analysis by Vogel et al (2006) reported mixed results in relation to gender, and advised caution when discussing gender differences. Second, in terms of ability, analysis of performance scores by levels of attainment provided two relatively clear pictures. There was a significant inverse relationship between ability and gains in scores; less competent children showed greater gains than those who were more able, with the most able showing the smallest gains. With speed of computation, the middle-ability children (level B, C and D) improved most. It may be that children who were of lower ability were concerned with improving the number of correct answers, whereas children who (presumably) had more confidence in their ability to compute accurately were more interested in improving their speed. Ceiling effects may have been a factor with the most able. Finally, it did not seem to make any difference whether the children had used the mental agility game at home. However, we suggest a degree of caution when interpreting this finding, since it was based on reports from the children; we were not in a position to verify claims of ownership, nor of engagement with the specific game.

Limitations
In principle, an area of uncertainty is treatment fidelity; no direct observations were conducted by the researchers. Nevertheless, several factors allow us to be relatively confident in relation to treatment fidelity. The console used (the Nintendo DS Lite) has a low skills threshold; the controls are easy to master, it requires no specialist technical knowledge and the programme is self-maintaining. In terms of the teacher’s daily routine, the opportunity cost is not high. It makes few, if any, demands on the teacher; indeed, feedback has led us to believe that it frees up time to undertake routine administration tasks at the start of the day.

In practice, some process data were available. Although the researchers had not specifically requested this, in several classes, children kept diaries in which they recorded their performance. In other schools, teachers had maintained an ongoing record of children’s performance on the
weekly 'Brain-Age Check'. In such cases, the data collected were examined by the researchers for significant deviations from agreed procedures. None were detected and at the conclusion of the project there were no reports from teachers of significant difficulties. On balance, while it is appropriate to acknowledge this limitation, we are reasonably confident that the game console was used as agreed at the outset.

Clearly, similar reservations apply in relation to the controls. Control teachers had been asked to continue with their normal practices during the treatment period. However, there were factors at work which might have made this difficult for some teachers. It was known that the project was formally supported by their employers and that it had also attracted the attention of the Inspectorate of Schools in Scotland (HMIE – HM Inspectorate of Education). This official interest, coupled with the fact that teachers knew their classes would be tested, may have resulted in the control teachers feeling under some pressure to ensure their classes performed well as a result of their ‘regular’ teaching. The fact that the controls showed significant gains in speed and accuracy of computation may indeed reflect the effects of time and effective teaching. But on the other hand, they may also be partially attributed to a John Henry effect; the teachers knew they were being compared with the experimental groups and so therefore ‘over-performed’.

The surprising finding that the improvements in performance—readily acknowledged by pupils in discussion—were not reflected in their mathematics self-concept raises issues of measuring self-perceptions. A detailed discussion of the complex field of self perceptions and their measurement is beyond the scope of this paper. However, given the possibility that children may not associate ‘games skills’ with school work, there may be value in looking beyond the narrow confines of academic-focused judgements of self (ie, mathematical self-concept, academic self-concept) in order to gauge whether children’s views of themselves have changed as a result of the new achievements. Recently, there has been an increased interest in the duality of self-esteem (Miller & Moran, 2006; Mruk, 1999; Tafarodi & Milne, 2002). One of the fundamental principles here is the belief in a generalised sense of competence as one of the two key components of self-esteem. This may be contrasted with the more specific and ‘academic’ judgements characteristic of self-concept methodology as adopted in the current study. An instrument which measures a generalised sense of competence may be more likely to pick up changes in children’s self-beliefs (possibly overcoming the difficulties associated with the domains of knowledge which have been discussed above) and any future work in this area might benefit from using a two-dimensional measure of self-esteem.

A final point here is that in any future replication of this work there will be value in retesting the children after a period of time, in order to learn whether performance is maintained.

**Implications and conclusions**

We believe that the nature of the research design allows a degree of confidence in the generalisability of the results. Nevertheless, several wider issues are raised by these findings. One concerns costs. A class set of game consoles involves a significant expense for a school. On the other hand, once bought, they could, with careful timetabling, be shared between many classes and used throughout the school in the course of a term. There is no guarantee that all children will prefer to practise tables and mental computation on a game console, although it would seem reasonable to suggest that most would. Certainly, our findings indicate that less-able children in particular may benefit from this type of activity. Although it was not investigated in the current study, the game consoles may hold a special attraction for those children least likely to want to do ‘routine’ schoolwork and hardest to motivate. One suspects that most teachers and managers understand the effect that this can have on the overall ethos in a class, and in a school.

There is a need to investigate whether this work has any ‘spin-off’ in other areas. In terms of children’s beliefs: do any increases in motivation—and confidence—transfer to other classroom
tasks? Do they influence attitudes to learning in other domains? At the level of cognitive functioning, it is clear that speed of mental calculation is central to the gains achieved in the current study; is it possible that this increased speed of processing information transfers to other learning tasks? Another area which would be interesting to investigate is the temporal dimension. What is the ideal time frame; is 9 or 10 weeks enough, or too much? Do the benefits persist over time? Will the children be able to maintain these gains and/or this positive attitude?

Many questions remain to be answered, but the current study adds to what has been learned from a small number of published studies into the use of COTS in classrooms. It is limited in the sense that it focuses on one narrow aspect of the curriculum only; it may tell us little about other aspects of classroom learning. It is important also to acknowledge that competence in mental arithmetic does not guarantee success in mathematics. Nevertheless, few would contest the view that basic numeracy is a fundamental building block for children in primary or elementary education. We believe that the design of the current study points towards high ecological validity and generalisability. Schools keen to harness the potential of games-based learning in this area may do so with some confidence.

Declaration of interests
No sponsorship was involved in this study. Funding was provided by Learning and Teaching Scotland, (a non-departmental public body funded by the Scottish Government) with an additional contribution from one of the participating Regional Education Authorities. The authors neither asked for, nor received, financial or any other form of support from Nintendo or any other commercial organisation.

References


### Appendix A: Self-concept instrument (based on Marsh, 1992)

Name: Class: Boy/Girl (please circle)

School: Date:  

<table>
<thead>
<tr>
<th></th>
<th>Mostly false</th>
<th>Sometimes false, sometimes true</th>
<th>Mostly true</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>False</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix B: Attitude to school scale
What do you feel about school?

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Mostly disagree</th>
<th>Sometimes disagree</th>
<th>Mostly agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Going to school each morning makes me happy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>The best days of the week are those at the weekend.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>School usually makes children bored.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Learning in school is necessary for future life.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>What children learn at school is worthless.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>