Gender differences in computerised and conventional educational tests

J. Horne
Department of Psychology, University of Hull, Hull, UK

Abstract
Research has demonstrated girls to outperform boys on conventional literacy tests. The present studies concern gender differences on computerised educational tests. Seventy-one children were tested using LASS Secondary and a set of seven conventional measures. No significant gender differences were found on any of the LASS Secondary modules, although females did outperform males on a conventional spelling test. A further 126 pupils were tested on computerised and paper versions of the LASS Secondary reading, spelling and reasoning modules. No gender differences were found on the computerised versions, but there were significant differences on the paper versions of the reading and spelling modules favouring females. In a third study, 45 children were administered computerised and paper versions of the LASS Junior reading and spelling modules. There were no significant differences on the computerised modules, but girls performed significantly higher than boys on the paper version of the spelling module. It is possible that computerised assessment does not detect the established gender effect due to differences between males and females in motivation, computer experience and competitiveness. Further large-scale studies are necessary to confirm these findings.

Keywords
assessment, children, computer, education, gender.

Introduction
Several researchers have shown girls to be scoring higher than boys on literacy tests on school entry at age 4 or 5 (Denton & West 2002; Justice et al. 2005) and even on pre-school cognitive tasks (Sammons et al. 2004; Sylva et al. 2004). Strand (1999) maintains that girls are typically found to score higher than boys in teacher ratings of reading, writing and mathematics throughout Key Stage 1. Indeed, a number of studies using baseline assessment schemes (mostly based on teacher rating scales) have shown girls to achieve higher scores than boys (Lindsay 1980; Strand 1997; Lindsay & Desforges 1999). In addition, the Annual Report on sample data from the QCA Baseline Assessment Scales for the Autumn 1998 and Spring and Summer 1999 periods (Qualifications and Curriculum Authority 1999) also shows that, from a nationally representative sample of 6953 children, girls score higher than boys on all items (which are based on teacher ratings). It has also been reported that girls make more progress than boys in reading and writing during Key Stage 1 and so increase the gender gap (Strand 1997, 1999). Strand (1999) reports that girls make less progress in mathematics but still score slightly higher than boys at age 7. However, Tizard et al. (1988) indicate that, on objective tests, there are no significant differences between boys and girls at the end of their nursery education (average age 4 years 7 months) but that girls are significantly ahead of boys in reading by age 7. Tymms (1999) reports significant gender differences, favouring girls, on all subsections of the PIPS baseline test. However, he argues that
‘none was large enough to be regarded as educationally important’ (p. 43). The computer-based version of the PIPS baseline test shows girls performing slightly better than boys, although the effect sizes (0.1–0.2) are small (CEM 2001).

Gender differences in education, particularly with girls scoring higher than boys on literacy, continue to be evident throughout the school years (Frome & Eccles 1998; Pajares et al. 1999; Swiatek et al. 2000; Herbert & Stipek 2005). A number of studies in recent years have demonstrated a female advantage over males on reading tests (MacMillan 2000; Diamond & Onwuegbuzie 2001) and writing tests (Malecki & Jewell 2003). Tizard et al. (1988) report that girls in Britain outperform boys on standardised tests of reading and writing by the age of 7 but there is only a small gender difference in mathematics. According to Fergusson and Horwood (1997) the traditional educational disadvantage of girls has been replaced by a male disadvantage. They studied a group of children from school entry to age 18 and found that males achieved less well than females. These differences were not due to differences in intelligence as boys and girls had similar IQ scores. Fergusson and Horwood suggest that the gender differences were explained by boys being more disruptive and inattentive in class, which impeded their learning. Male (1994) states that, in one LEA, two-thirds of the children receiving a statement of special educational needs are boys. There are also a disproportionate number of boys identified as having reading difficulties in the United States (Allred 1990). Furthermore, Vardill (1996) reports that more boys than girls are identified as having special educational needs and, in one LEA, almost twice as many boys, compared with girls, were referred. He suggests that one of the reasons for this is that boys experiencing difficulties are more likely to be a problem to the teacher because of associated conduct problems. Alternatively, Malecki and Jewell (2003) suggest that boys may be over-identified for difficulties as girls have an advantage on tests and normative data may not account for gender.

Plewis (1997) suggests that the reasons for gender differences in educational attainment are not well understood. It has sometimes been argued that teachers have lower expectations for boys. He suggests, however, that to conclude that teachers are biased against boys it is necessary to show that these low expectations are inaccurate and that teachers behave differently towards male pupils. Plewis did not find teachers’ assessments to be less accurate for this group, although he did find teachers’ expectations to be biased against boys. A number of researchers suggest that expectations could be affected by the way boys behave (Tizard et al. 1988; Bennett et al. 1993; Delap 1995).

Despite the huge growth in expenditure on information and communication technology in education over recent years, the majority of assessment taking place in schools is pen-and-paper based (Cowan & Morrison 2001). According to Hargreaves et al. (2004), English schools have one of the highest rates of computer use in the world, with one computer for every 9.7 pupils in 2002. There are a number of advantages of computerised assessment (see Greenwood & Rieth 1994; Singleton 1994, 1997; Singleton et al. 1995, 1999; Woodward & Rieth 1997; British Psychological Society 1999) including that they are labour and time saving, more standardised and controlled. Another principal advantage is the possibility of adaptive testing. In an adaptive test, individuals are moved quickly to the test items that will most efficiently discriminate his or her capabilities, making assessment shorter, more reliable, more efficient and often more acceptable to the person being tested. A number of studies have indicated that children (particularly if they feel they might perform badly) often prefer computer-based assessment to traditional assessment by a human assessor. However, the effect of utilising computerised assessments on gender differences is not clear. It has been suggested that females may perform more poorly than males on computerised tests due to computer anxiety (Brosnan 1999; Todman 2000) although it has also been argued that males under-perform on such tests as they practice less. A number of studies have shown females to obtain lower scores on computerised assessments (Gallagher et al. 2002; Volman et al. 2005), while others have shown there to be no gender differences on computerised assessment (Singleton 2001; Fill & Brailsford 2005).

The present study seeks to ascertain whether gender differences exist in computerised educational tests. It consists of three small-scale studies involving two computerised tests – one for secondary schools (LASS Secondary) and one for junior schools (LASS Junior).
Method

Participants

The participants in the three studies attended schools in different regions of England and Wales, which were selected in order to give a representative range across the socio-economic spectrum. Details of the participants are given in Table 1. The pupils in each school were randomly selected from the school register – they were not selected on the basis of ability or for being considered as at risk of having any special educational need.

Instruments

Study 1: LASS Secondary versus conventional tests

LASS Secondary (Horne et al. 1999) is a computerised test consisting of seven main assessment modules. It is standardised for the age range 11 years 0 months to 15 years 11 months.

1. Sentence reading is an adaptive test that involves finding the missing word in a sentence. Items are selected from a bank of 70 items.
2. Spelling is an adaptive test that involves spelling single words. Items are selected from a bank of 98 items.
3. Reasoning is an adaptive test involving logical matrix puzzles. Items are selected from a bank of 58 items.
4. Visual spatial short-term memory has a scenario of a ‘cave’ with eight ‘hollows’. Different ‘phantoms’ appear in different hollows one at a time and then disappear. The pupils must select from an array the phantoms that were presented and place them in the correct hollow. This is an adaptive test that starts with three phantoms and progresses to a maximum of eight phantoms, depending on the pupils’ performance.
5. Auditory-sequential short-term memory is a forward digit span task. Pupils are given a telephone number to remember which they then enter on to a representation of a mobile phone using the mouse or the computer keyboard. Pupils start with two items of three-digit numbers and, if they answer one or both correctly, then they move on to two items of four-digit numbers and so on up to a maximum of nine digits.
6. Non-word reading is a test of phonic decoding skills and has 25 items. For each item, pupils are presented with a non-word visually and hear four different spoken versions of the word; they must select the version that they think is correct.
7. Syllable segmentation is a test of syllable and phoneme deletion, which identifies poor phonological processing ability. Pupils are presented with 32 words and asked what each word would sound like if part of it is removed. They listen to four alternative answers, before selecting the answer that they think is correct.

Commensurate conventional tests were used as comparison measures for the seven LASS Secondary modules (see Table 2). It is important to note that these tests are not exact equivalents of the LASS Secondary modules.

Study 2: LASS Secondary computerised versus pen-and-paper administration

The second study utilised the LASS Secondary (Horne et al. 1999) sentence reading, spelling and reasoning modules as outlined above. These were administered in their standard computerised format (as adaptive tests) and also in a pen-and-paper format (in which all items were administered).

Study 3: LASS Junior computerised versus pen-and-paper administration

The final study utilised the LASS Junior (Thomas et al. 2001) sentence reading and spelling modules. These were administered in their standard computerised format.
format (as adaptive tests) and also in a pen-and-paper format (in which all items were administered).

1. Sentence reading is an adaptive test that involves finding the missing word in a sentence. Items are selected from a bank of 94 items.
2. Spelling is an adaptive test that involves spelling single words. Items are selected from a bank of 122 items.

Procedure

Study 1: LASS Secondary versus conventional tests
Pupils completed the seven LASS Secondary modules, with each school completing the tests in a different order. Pupils were administered the equivalent conventional tests, in the same order, with a 4-week gap in between. Half the schools completed the LASS Secondary tests first and half completed the conventional tests first.

Study 2: LASS Secondary computerised versus pen-and-paper administration
Pupils completed the three LASS Secondary computerised modules, with each school completing the tests in a different order. Pupils were administered the equivalent pen-and-paper versions, in the same order as the comparable computerised tests, with an 8-week gap in between.

Study 3: LASS Junior computerised versus pen-and-paper administration
Pupils completed the two LASS Junior computerised modules and the equivalent pen-and-paper versions, with a 4-week gap in between. Half the pupils completed the computerised tests first and half completed the pen-and-paper tests first.

Results

Study 1: LASS Secondary versus conventional tests
Analysis by gender (see Table 3) revealed no significant differences on any of the LASS Secondary modules.

Analysis by gender revealed a significant difference on only one of the conventional measures (BSTS 3 spelling test), with girls outperforming boys on this test (see Table 4).

Fifty-one of the 71 pupils preferred the computer tests, while 17 preferred the paper-and-pencil tests. There was no significant gender difference (see Fig 1) in the preference for computerised or paper-and-pencil tests ($\chi^2 = 0.34, df = 1$, not significant).

Study 2: LASS Secondary computerised versus pen-and-paper administration
Analysis by gender (see Table 5) revealed no significant differences on any of the LASS Secondary computerised modules but significant differences on two of the pen-and-paper versions of the tests (reading and spelling), with girls outperforming boys on these tests. Correlations between the computerised and pen-and-paper versions of the tests were 0.90 ($P < 0.001$) for reading, 0.89 ($P < 0.001$) for spelling and 0.79 ($P < 0.001$) for reasoning.

Study 3: LASS Junior computerised versus pen-and-paper administration
Analysis by gender (see Table 6) revealed no significant differences on either of the LASS Junior
computerised modules but a significant difference on the pen-and-paper version of the spelling test, with girls outperforming boys on this test. Correlations between the computerised and pen-and-paper versions of the tests were 0.92 ($P < 0.001$) for reading and 0.96 ($P < 0.001$) for spelling.

**Discussion**

No significant gender differences were found on any of the computerised LASS Secondary or LASS Junior modules. The conventional measures utilised in Study 1 show a significant gender difference on spelling, with females outperforming males. Additionally, girls also scored higher than boys on the pen-and-paper versions of the LASS Secondary reading and spelling modules and the LASS Junior spelling module. The gender difference in spelling ability on the three pen-and-paper spelling tests is concordant with findings by Hyde and Linn (1988) that, even at an undergraduate level (Coren 1989), a gender difference, favouring females, exists in spelling. The results of the first LASS Secondary study suggest that there is no significant effect of gender in pupils’ preference for computerised or conventional tests.

The findings of the present study, that there are no significant gender differences in the computerised reading and spelling tests, are in contrast to the results...
of previous studies, which suggested that girls out-perform boys in conventional pen-and-paper literacy tests. It could be suggested that fully computerised tests are more objective than conventional assessment and less susceptible to gender bias. However, it is more likely that the interaction with a computer in the studies has enhanced the motivation of the boys more than that of the girls and this could have compensated for (or masked) gender differences in performance. Several studies have shown that boys globally use ICT more both at home and at school (Scott et al. 1992; Bannert & Arbinger 1996; Makrakis & Sawada 1996; Janssen & Plomp 1997; Kadijevich 2000). Further, it is reported that girls do not view computers as positively as boys do (Durdell 1991; Comber et al. 1997; Durdell & Thomson 1997; Huber & Schofield 1998) and Crook (1996) suggests that girls’ attitudes towards technology may become more negative as they go through school. Previous research has suggested that females experience higher levels of computer anxiety (Brosnan 1999; Todman 2000) which may negatively affect their performance on computerised tests. This is consistent with the finding of Newton and Beck (1993) that, in the early 1990s, the percentage of women studying for computer science degrees was falling. Additionally, Volman and van Eck (2001) state that

![Graph showing gender differences in preferences for different test types.](image)

**Fig 1.** Gender differences in preferences for different test types.

### Table 5. Gender differences on the LASS Secondary computerised and paper-administered modules.

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>df</th>
<th>Significance</th>
<th>Effect size (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computerised reading</td>
<td>Female</td>
<td>66</td>
<td>53.95</td>
<td>18.93</td>
<td>0.21</td>
<td>113</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>57</td>
<td>53.19</td>
<td>21.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computerised spelling</td>
<td>Female</td>
<td>65</td>
<td>59.85</td>
<td>20.64</td>
<td>0.90</td>
<td>121</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>58</td>
<td>56.55</td>
<td>19.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computerised reasoning</td>
<td>Female</td>
<td>51</td>
<td>33.90</td>
<td>17.76</td>
<td>0.69</td>
<td>93</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>46</td>
<td>36.13</td>
<td>14.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pen-and-paper reading</td>
<td>Female</td>
<td>66</td>
<td>55.20</td>
<td>16.45</td>
<td>3.14</td>
<td>121</td>
<td>P&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>57</td>
<td>45.32</td>
<td>18.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pen-and-paper spelling</td>
<td>Female</td>
<td>65</td>
<td>61.05</td>
<td>20.21</td>
<td>3.37</td>
<td>121</td>
<td>P&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>58</td>
<td>48.78</td>
<td>20.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pen-and-paper reasoning</td>
<td>Female</td>
<td>51</td>
<td>35.63</td>
<td>13.48</td>
<td>0.16</td>
<td>95</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>46</td>
<td>35.24</td>
<td>10.04</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 6. Gender differences on the LASS Junior computerised and paper-administered modules.

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>df</th>
<th>Significance</th>
<th>Effect size (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computerised reading</td>
<td>Female</td>
<td>20</td>
<td>64.05</td>
<td>18.66</td>
<td>1.16</td>
<td>43</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>25</td>
<td>56.48</td>
<td>24.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computerised spelling</td>
<td>Female</td>
<td>20</td>
<td>75.20</td>
<td>19.76</td>
<td>1.60</td>
<td>43</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>25</td>
<td>65.00</td>
<td>22.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pen-and-paper reading</td>
<td>Female</td>
<td>20</td>
<td>61.85</td>
<td>15.43</td>
<td>1.43</td>
<td>43</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>25</td>
<td>54.12</td>
<td>19.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pen-and-paper spelling</td>
<td>Female</td>
<td>20</td>
<td>79.25</td>
<td>18.14</td>
<td>2.19</td>
<td>43</td>
<td>P&lt;0.05</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>25</td>
<td>66.12</td>
<td>21.30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
girls report fewer ICT skills and less ICT knowledge than boys and Janssen and Plomp (1997) suggest that boys score better than girls on tests of ICT skills. However, Crook and Steele (1987) found no significant gender differences among reception class pupils choosing to engage in computer activities in school and Essa (1987) obtained similar results among pre-schoolers. Volman and van Eck (2001) and Volman et al. (2005) also found gender differences to be less common in younger pupils than older pupils. Crook (1996, p. 25) concludes that ‘...gender-based attitude differences are not convincingly present at the start of schooling: they must somehow be cultivated within the early school years’. Nevertheless, if gender differences in computer interest do exist, then these could bias the results of a computerised assessment. However, Taylor et al. (1999) found no relationship between computer familiarity and level of performance on a computerised test of English as a foreign language, after controlling for English language ability. Furthermore, Hargreaves et al. (2004) found no relationship between low performance on a computerised mathematics test and unfamiliarity or nervousness with computers. This has been supported in adult samples by Parshall and Kromrey (1993) and Clariana and Wallace (2002). Fill and Brailsford (2005) found no evidence of higher computer anxiety among female University students.

Alternatively, it is suggested that males are more competitive than females, while females are more collaborative (Maccoby & Jacklin 1974; Fiore 1999). According to Hayes and Waller (1994) this may interfere with accurate performance under situations such as group testing. The pen-and-paper reading and spelling tests used in the studies were administered as group tests, while the computerised tests were administered individually in most cases. It is therefore possible that the male students attempted to finish the conventional tests quickly and so made more mistakes, whereas during the individual computerised tests boys were under less pressure to complete the test quickly. Both LASS Junior and LASS Secondary have network versions, which can be installed on several machines within a computer room so that pupils can be tested simultaneously (wearing headphones). In such situations, it may be a necessary precaution to advise that students being tested at the same time be administered the tests in different orders. Furthermore, the paper versions of the LASS tests in the current studies incorporated all of the test items, while the computerised versions utilised an adaptive algorithm. It is possible that the extended length of the paper test impacted more negatively on the boys.

Conclusions

Gender differences favouring females that are evident in conventional literacy tests are not evident in the LASS Secondary and LASS Junior reading and spelling tests. It may be that computerised assessment does not detect this established gender effect due to differences between males and females in motivation, computer experience and competitiveness. However, these results are from small-scale studies and further large-scale studies looking at the effect of test mode on gender differences in educational tests are necessary to establish the generalisability of these findings.

References


