Journal of Research in Reading

UKLA The United Kingdom Literacy Association

Journal of Research in Reading, ISSN 0141-0423 *DOI: 10.1111/j.1467-9817.2009.01402.x*
Volume 32, Issue 4, 2009, pp 402–412

Typography for children may be inappropriately designed

Arnold Wilkins, Roanna Cleave, Nicola Grayson and Louise Wilson

Department of Psychology, University of Essex, Colchester, UK

We present four studies indicating that the size and design of the typeface in textual material for children aged 7–9 may impair speed of reading and comprehension, and measurement of reading attainment. The first study compared the speed with which sample sentences were comprehended. The sentences were printed in Arial font with an x-height of 4.2 or 5.0 mm. The sentences were verified 9% more quickly when presented in the larger typeface. The second study compared reading age on the Salford Sentence Reading Test when the typeface remained at the initial size (x-height 3.3 mm) throughout the test, and when it decreased in size as usual. The average reading age measured with the larger font was 4 months older. The final studies compared the font Sassoon Primary with the font Verdana and showed that Verdana was read and searched more quickly.

We present four studies indicating that the size and design of the typeface in textual material for children may be suboptimal, impairing speed of reading and comprehension, and measurement of reading attainment.

The various typographic parameters of font size, inter-character spacing, word spacing, line spacing, justification and line length interact in affecting reading performance (e.g. Tinker, 1968).

In children's reading material there is additional complexity. The shape of characters may differ from those in adult text, particularly as regards single-storey 'a' and 'g' (infant letters). Walker and Reynolds (2003, 2004) have argued that the two-storey adult forms of a and g are familiar to children who have begun to read and may be preferred to the infant forms because they are less confusable.

Most typefaces for children are sans serif. Visual psychophysical studies in adult readers have shown that serifs have little effect on legibility (Arditi & Cho, 2005), and studies in children concur (Walker & Reynolds, 2003, 2004).

In children's text the size and spacing of characters is usually large initially and decreases with reading age. The x-height (height of the central body of letters) is typically about 4 mm in children's early readers and decreases to adult size ($\sim$ 2 mm) over the course of 5 years. Hughes and Wilkins (2000) have argued that the typeface becomes too small too early in life. They asked children aged 5–11 to read aloud paragraphs of randomly ordered common words for a period of 1 minute, and measured the number of

words correctly read. The passages were printed in a typeface and type size similar to that in two widely used reading schemes for children. The x-height ranged from about 4 mm (typical for 5-year-olds) to 2 mm (typical for 11-year-olds). The same 15 high-frequency words, familiar to all participants, were used for all passages, regardless of size, which meant that the effect of typeface could be investigated independently of word difficulty. The words were arranged in random order, 15 to a line, as a paragraph of text. The children were unable to guess a word from context but had to see it in order to read it. Because the passages were meaningless, the children were usually unaware of any errors of omission and transposition. Reading speed is the best single measure of performance in this task because rate and accuracy are positively correlated: there is no speed/accuracy trade-off (Hughes & Wilkins, 2000; Wilkins, Jeanes, Pumfrey & Laskier, 1996). The study showed that children aged <10 years read text of a size designed for younger children more quickly than text of a size that they were familiar with.

The reading task used by Hughes and Wilkins (2000) was unlike normal reading because the material was meaningless and did not involve comprehension. The purpose of the first of four studies reported in this paper was to rectify this shortcoming and investigate the effect of font size on silent reading for comprehension. Most tests of reading comprehension (e.g. Neale Analysis of Reading Ability; Neale, 1999) involve asking questions about the content of a passage once the passage has been read. Performance on this test involves memory as well as comprehension and is not very sensitive to the speed of comprehension, a function that might be influenced by the speed with which text can be read. The Speed and Capacity of Language-Processing Test (SCOLP; Baddeley, Emslie & Nimmo-Smith, 1992) is different. It requires individuals to determine the veracity of simple sentences, based on their knowledge of the world. The child is required to place a tick in a box beside sentences such as 'Bicycles have wheels' and a cross beside sentences such as 'Flowers wear clothes'.

The SCOLP was used in the first study to examine the speed with which such sentences could be comprehended when presented in typefaces of different size. The sentences were presented in a font familiar to 7-year-old children but in a font size similar to that used with 5-year-olds, and also in a size 19% larger.

Study 1: comprehension speed

Participants

A class of children in Year 3 in a mainstream county primary school in Suffolk took part: 10 boys and 14 girls, aged 7:3–8:3. Three children with reading ages below 5 (as evaluated by the Salford Sentence Reading Test [Revised]; Bookbinder, Vincent & Crumpler, 2002) were excluded. Those children who habitually wore glasses used them for the study.

Materials

The SCOLP was prepared using Arial font. The two versions of the test, A and B, were printed in both large (26 pt) and small (22 pt) font. The large font had an x-height (height of the central body of the letters) of 5.0 mm and the small had an x-height of 4.2 mm. Figure 1 shows an example of the material from the two versions of the test. The x-height of 4.2 mm was similar to that conventionally used with 5-year-olds (Hughes & Wilkins, 2000).

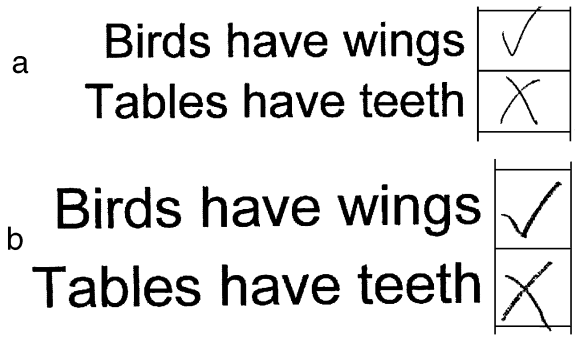

Figure 1. Fragments of the Speed and Capacity of Language-Processing Test (SCOLP) sentence comprehension test: (a) 22 pt; (b) 26 pt.

Procedure

The children were tested individually. Half the participants (Group 1) used Version A with large text and Version B with small, and half (Group 2) experienced the opposite combination. Half the participants in each of these groups experienced Version A first, and half Version B. They were asked to place a tick against correct sentences such as 'Fish live under the water' and a cross against incorrect sentences such as 'People have two noses'. They were given these sentences as examples at the outset, and three sentences to complete under supervision. They then completed 20 sentences as quickly as they could. After a break they were given the second version to complete, a further 20 sentences.

Results

An average of 17.1/20 sentences for the small font and 17.3/20 for the large font were categorised correctly as true or false. The difference in error rates was not significant. The mean time to complete 20 sentences was 212 seconds for the small font, and 193 seconds for the large, a 9% difference in speed that was statistically highly significant, $t(23) = 3.69$, $p < .001$. There was no significant effect of version or order.

Discussion

Sentences presented in a font of a size larger than is typical for use in material for 5-year-olds were comprehended by 7–8-year-olds more rapidly than those of a more conventional size. The difference in size approximated 19% and it resulted in an increase in reading speed of 9%. The increase in reading speed was accompanied by a marginal increase in accuracy, indicating that the increase in speed was not attributable to a greater tolerance of inaccuracy: there was no speed/accuracy trade-off.

In many reading tests the font size gets smaller for those passages designed to be read by children with higher reading age. The purpose of the second study was to investigate the effect of font size on reading age measured using such tests.

Study 2: reading age

Reading attainment tests used in schools usually present material that increases progressively in difficulty. Children are asked to read the material until it becomes too difficult for them to read accurately. As the material increases in difficulty, so the size of font in which it is printed decreases. One example is the Salford Sentence Reading Test (Revised) (Bookbinder et al., 2002). Sentences are presented in Sassoon Primary, initially with an x-height (height of the central body of the letters) of 3.3 mm, and then 2.4, 2.2 and 1.6 mm.[1]

Participants

A class of children in Year 3 of a county primary school in Essex participated: 18 boys and 29 girls, aged 8:1–9:6. Those children who habitually wore glasses used them for the study.

Materials

The two versions of the Salford Sentence Reading Test (Revised) were each prepared in the conventional edition in which the typeface decreased in size with sentence complexity, and an alternative edition in which the size of the type used with the initial sentences (24 pt, x-height 3.3 mm) was used for all sentences of the test. The font used was Sassoon Primary. Figure 2 shows an example of the material.

Procedure

The children were tested individually. Half the children received the conventional edition in one version and the constant text-size edition in the other version, and half received the opposite pairing. Half of each group undertook the conventional edition before the constant size edition, and half the reverse order, so that the presentations were counterbalanced for order and version. The child was introduced to the test in the style recommended in the Salford Sentence Reading Test (Revised) handbook: 'I would like you to read some sentences for me. They are easy at first but each one gets a little harder'. As per the handbook, an error was recorded if the child (1) took longer than 6–7 seconds to read a word, (2) read a word incorrectly or (3) skipped a word. An error was not recorded if the child made a spontaneous correction. Once six errors had been recorded the test ended. The second version of the test was then presented in the same way.

Results

Fifteen children were at the test ceiling (10 years 3 months) on both editions of the Salford Sentence Reading Test (Revised) and their data were excluded from further analysis. The chronological age of these children averaged 8 years 9 months (range 8:1–9:6). For the remaining 32 children (mean age 8:9, range 8:1–9:6) the reading age

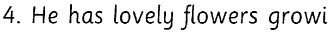

Figure 2. A fragment of material from the Salford Sentence Reading Test (Revised).

measured on the conventional edition of the test was 7:8, and the reading age (of the same children) measured on the constant size edition was 8:1. The difference of a little over 4 months was equivalent to about 25% of the deviation of the scores for the sample, and was significant on a 2×2 mixed design analysis of variance, $F(1, 30) = 5.41$, $p < .05$. There was no significant effect of test order, $F(1, 30) = 0.06$, $p = .80$. Only three children were at ceiling on one version of the test (that with constant text size) and not on the other. When these individuals were removed the difference in reading age was not substantially affected. Forty-one per cent of children achieved a higher reading age than chronological age on both versions of the test.

Discussion

The small size of the text the children were required to read in the conventional version of the reading test reduced reading performance. The level of scholastic attainment in reading measured by the conventional version of the test in which text decreased in size with increasing age was about 4 months lower than the reading age measured when the same test was printed in large text throughout. Note that the largest font size had an x-height (height of the central body of the letters) of 3.3 mm, and was therefore smaller than the smaller of the two font sizes used in the previous study. Had the font been larger, the reading age, as measured, might have been even higher.

The font used in this test was Sassoon Primary, designed to have letter shapes familiar to children. This font has been shown to have a high similarity between neighbouring letter strokes, and such similarity impairs reading. In the next study, it is shown that this font is less easy to read than a more conventional font.

Study 3: font design

Wilkins et al. (2007) showed that the similarity in shape between the neighbouring strokes of letters within a word made some words appear striped. The inter-stroke similarity predicted not only the striped appearance of a word but also the speed with which it could be read by adult fluent readers and by children. The similarity in shape between neighbouring strokes of letters was measured using the mathematical technique of autocorrelation. The technique can be understood as follows. Imagine two identical copies of the image of the word superimposed. One of the copies is displaced horizontally by one pixel and the number of superimposed black pixels in the two images is counted. The horizontal displacement is repeated, pixel by pixel, and the number of superimposed black pixels plotted as a function of displacement. The number initially decreases with displacement, but as one letter stroke is superimposed on its neighbour the number of coincident pixels increases so as to reach a peak before decreasing once again. The variation in number of coincident black pixels with displacement provides a function similar to the horizontal autocorrelation function. The higher the initial peak, the greater the similarity in shape between neighbouring strokes of letters within the word, and the greater their periodicity.

Wilkins et al. (2007) created passages of randomly ordered common words with high initial peak and similar passages of words with low peak. The former were read aloud about 5% more slowly than the latter by fluent adult readers. Visual search for specific

words in a passage of randomly ordered words was 40% slower in the passages of words with high first peak than in those with words of low.

The first peak in the autocorrelation was larger for some fonts than for others, averaged over a corpus of 1,000 words. Generally, fonts with the larger first peak were read more slowly. Times New Roman and Sassoon Primary were among the fonts with high first peak, and Verdana and Tahoma among those with lower first peak. The purpose of the third and fourth studies was to see whether Sassoon Primary, used in many primary schools in the United Kingdom, is read more slowly than an alternative adult font, Verdana, which has lower similarity in shape between neighbouring letter strokes. Study 3 measured the speed of reading aloud and Study 4 the speed of silent visual search.

Participants

Two classes from Year 4 in a mainstream county primary school took part: 38 boys and 42 girls, aged 8:2–9:2. Twenty wore glasses for reading and for the study.

Materials

The Rate of Reading Test (Wilkins et al., 1996) consists of 15 randomly ordered common words arranged in a paragraph of text, the same words in different random order on each line (e.g. come see the play look up is cat not my cat and dog for you to). The words are of high frequency and familiar to poor readers. Each passage is read aloud as quickly as possible for 1 minute, and the score is the number of words correctly read. The score has been shown to be highly reliable at re-test, and to predict the benefits of coloured filters (Wilkins & Lewis, 1999). Two passages of the Rate of Reading Test were prepared in Sassoon Primary 13.5 pt and 11.0 pt Verdana. The two fonts, Sassoon and Verdana, were matched in x-height (2.25 mm). The passages generated in Microsoft Word had 'single' line spacing. As a result, material presented in Sassoon had an interlinear spacing of 4.7 mm and the material set in Verdana had an interlinear spacing of 4.3 mm. The text body had a width of 116 mm for all passages. Figure 3 shows an example of the passages. Note that the adult two-storey form of the a was used, and that Verdana uses a single-storey g.

Procedure

The children were tested individually. At the outset of testing the children were asked to read aloud the 15 words from the Rate of Reading Test presented as a vertical list in Times New Roman 14 pt. All children could read them correctly without prompting.

The children were then given a practice passage of the Rate of Reading Test printed in Times New Roman 14 pt to avoid any differential range effects concerning the fonts under study. They were then given the passages printed in Sassoon and Verdana in an order counterbalanced for font and passage. They were asked to read each passage aloud

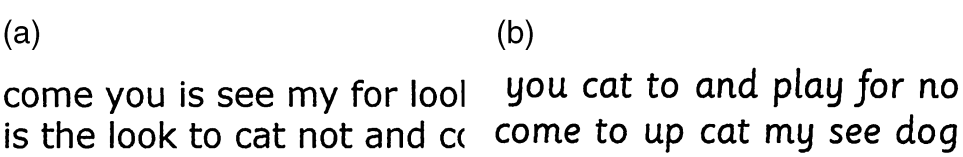

Figure 3. Fragments of the material used in Study 3: (a) Verdana; (b) Sassoon Primary.

as quickly and as accurately as possible, and were given 1 minute per passage in which to do so. They were then asked which passage they preferred.

The passages were scored in terms of the number of words correctly read. A word had to be in the correct order relative to its neighbours to count as correct. Only one word in a transposed pair therefore counted as correct.

Results

The mean number of words correctly read was 87.2 (*SD* 23.7) for the passages printed in Sassoon and 92.1 (*SD* 21.9) for those printed in Verdana. The difference was highly significant, $t(79) = 3.09$, $p < .01$, two-tail. There was no effect of test order. There was no significant difference in accuracy between the two fonts. Sixty per cent of the children expressed a preference for Verdana ($p = .07$ relative to chance expectation). Their preference was not significantly related to performance.

The task of reading randomly ordered common words aloud at speed, as used in this study, has been shown to be sensitive to visual aspects of reading material, such as its colour (Wilkins, Lewis, Smith & Rowland, 2001) but it is not a task that resembles skilled reading. For this reason, the next study investigated the speed with which children could search silently for specific words and identify their location in a passage.

Study 4

Participants

One of the classes of children who had participated in the previous study was available to take part: 22 girls and 21 boys, aged 8:3–9:3. Eight wore glasses for reading and for the study.

Materials and procedure

The passages from the previous study were used: the practice passage in Times New Roman, and two passages in Sassoon and two in Verdana. Children were tested individually. All children experienced the four passages in an order that ensured that all possible combinations of word sequence, font and target location (positioned at random within the first or second half of the passage) were balanced across participants. Children were instructed to search for a pair of consecutive words spoken by the examiner, and the time they took to find them was recorded.

Results

Two children could not find the words in one or more passages and their data were excluded. The speed of visual search was calculated by dividing the time taken to find a pair of target words by the number of preceding words in the passage. The visual search speed (estimated from the mean search time) in Sassoon Primary was 1.72 words per second and for Verdana 2.46 words per second, a difference of 43%. The difference was significant, $t(40) = 1.72$, $p = .034$, one-tail. The above analysis assumed a serial search, and such an assumption may be unwarranted. An alternative analysis was therefore

undertaken based simply on the time taken to find a target regardless of its position. The average time taken for the passages in Sassoon was 49 seconds and for those in Verdana 34 seconds, a difference of 45%. A repeated measures analysis of variance showed the main effect of font to be significant, $F(1, 40) = 7.3$, $p = .01$ and there was no significant effect of order, $F(1, 40) = 2.3$, $p = .13$.

Discussion

Two commonly used fonts have been compared: one with a high similarity of shape and regularity between neighbouring strokes of letters, and one with a lower similarity of shape and regularity. As in previous studies (Wilkins et al., 2007), the font with the greater similarity of shape/regularity was read and searched more slowly. Sassoon Primary is a font used in primary schools because the letter shapes resemble those that the children are familiar with when they begin to read. Verdana is not commonly used in this context: the ascenders and descenders are thought to be too short, making discrimination of letter shape difficult (Walker & Reynolds, 2003, 2004). The present results demonstrate that letter shape needs to be studied in the context of the word, and that the internal word shape, as expressed in the horizontal autocorrelation, may be important. The children read more slowly with Sassoon than Verdana both aloud and silently, even though Sassoon was a font with which they were more familiar and even though the spacing of the lines of text was larger.

General discussion

It may be helpful to consider the visual demands faced by a child learning to read. Unfamiliar shapes need to be visually decoded. The unfamiliar shapes are similar to one another and their significance can often only be appreciated by virtue of their relative position. Relative position is confusable, because similar shapes occur in many positions. The position of a letter shape can be determined by its context within the string of letters that comprise a word, but only if the word is familiar. As the child develops and becomes more familiar with written language, the visual demands of the task recede: top-down processes of word recognition can be brought to bear on the visual clutter of confusingly similar shapes. But until familiarity with words is achieved, the task of reading may place stringent demands on the visual system. The demands on vision can be appreciated by inspection of Figure 4, which shows a passage of Thai reproduced in large and small font. The symbols will be unfamiliar to most readers of this article. The familiarity of symbols and strings of symbols cannot be used to determine the relative position of shapes, information that is required in order to read. The visual demands imposed by reading unfamiliar material should be reduced if letter shapes are as different as possible. They should also be reduced if the letters are of a size sufficient that their position is readily determined. Compare Figure 4a and b.

A large font with relatively few words on the page improves information about the position of a symbol upon the page.

The benefit of a large font (and larger associated spacing) has now been demonstrated in a range of tasks. Hughes and Wilkins (2000) used meaningless material that was read aloud and showed that reading speed increased with a font larger than that typically used.

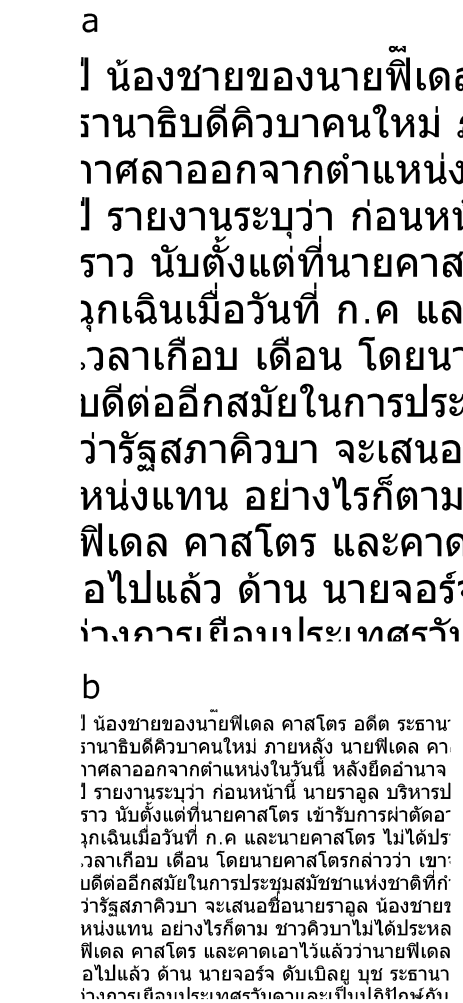

Figure 4. Fragment of a passage of Thai to illustrate the visual effects of unfamiliar crowded letters.

We have extended their results and shown in Study 1 that the typeface in most reading schemes is typically too small for the rapid silent comprehension of simple meaningful sentences. The size of typeface also has consequences for the assessment of reading skill. Study 2 showed that performance on a conventional test of reading ability, requiring accurate reading of sentences, was impaired by the font size it used.

As mentioned in the introduction, typographical parameters cannot be considered in isolation; they interact in affecting visual performance. In Studies 1 and 2 the increase in font size was necessarily associated with a proportionate increase in line length and spacing. In Studies 3 and 4, x-height was controlled, but line spacing was larger for the Sassoon Primary font. The passages in Sassoon Primary were read more slowly despite the larger line spacing. The present studies therefore go some way towards demonstrating that reading ability and comprehension may be adversely affected by current typographic practice, but it remains to be discovered how best to improve upon such practice within practical constraints.

The determination of the position of symbols within a symbol string is likely to be made more difficult if the shapes of the symbols are similar one to another, and this varies with typeface design. Sassoon and Verdana were selected for comparison not because they are representative of the fonts used in schools, but because they represent extremes of the differences in the similarity/regularity of neighbouring letter strokes encountered in common fonts. The findings suggest that, in principle, the similarity of internal word shape expressed in these terms affects both reading speed and visual search.

Many children report perceptual distortions of text, illusions of motion shape and colour. The distortions are thought to occur because of the striped properties of text, both those from the letter strokes (Wilkins et al., 2007) and the lines of text (Wilkins, 1995, chapter 5). Children who experience distortions tend to be particularly affected by the size of typeface (Hughes & Wilkins, 2000) and it is these individuals who benefit from coloured overlays (Hollis & Allen, 2006). About 5% of children in mainstream education read at least 25% more quickly with a coloured overlay (Wilkins et al., 2001). If the text they are required to read were less stressful, and the classrooms were not overlit with flickering fluorescent lighting (Winterbottom & Wilkins, 2009), fewer such aids would doubtless be required.

Acknowledgements

We thank the children and staff at Claydon Primary School and Wells Hall Community Primary School, Suffolk, and the other participating primary schools.

Note

1. Sassoon Primary was not designed for use by nine-year-old children (R. Sassoon, Personal communication).

References

Arditi, A. & Cho, J. (2005). Serifs and font legibility. *Vision Research*, 45, 2926–2933.

Baddeley, A.D., Emslie, H. & Nimmo-Smith, M.I. (1992). *The speed and capacity of language processing task: SCOLP*. Bury St Edmunds: Thames Valley Test Company.

Bookbinder, G.E., Vincent, D. & Crumpler, M. (2002). *Salford sentence reading test (revised)*. London: Hodder Education.

Hollis, J. & Allen, P.M. (2006). Screening for Meares–Irlen sensitivity in adults: Can assessment methods predict changes in reading speed? *Ophthalmic and Physiological Optics*, 26(6), 566–571.

Hughes, L.E. & Wilkins, A.J. (2000). Typography in children's reading schemes may be suboptimal: Evidence from measures of reading rate. *Journal of Research in Reading*, 23(3), 314–324.

Neale, M. (1999). *Neale analysis of reading ability*. (3rd edn, Australian Standardisation). Camberwell, Vic., Australia: ACER Press.

Tinker, M.A. (1968). Suitable typography for beginners in reading. *Education*, 88(4), 317–320.

Walker, S. & Reynolds, L. (2003). Serifs, sans serifs and infant characters in children's reading books. *Information Design Journal*, 11(2/3), 106–122.

Walker, S. & Reynolds, L. (2004). 'You can't see what the words say': Word spacing and letter spacing in children's reading books. *Journal of Research in Reading*, 27, 87–98.

Wilkins, A.J. (1995). *Visual stress*. Oxford: Oxford University Press.

Wilkins, A.J., Jeanes, R.J., Pumfrey, P.D. & Laskier, M. (1996). Rate of reading test: Its reliability, and its validity in the assessment of the effects of coloured overlays. *Ophthalmic and Physiological Optics*, 16, 491–497.

Wilkins, A.J. & Lewis, E. (1999). Coloured overlays, text and texture. *Perception*, 28, 641–650.

Wilkins, A.J., Lewis, E., Smith, F. & Rowland, E. (2001). Coloured overlays and their benefits for reading. *Journal of Research in Reading*, 24(1), 41–64.
Wilkins, A.J., Smith, J., Willison, C.K., Beare, T., Boyd, A., Hardy, G. et al. (2007). Stripes within words affect reading. *Perception*, 36(12), 1788–1803.
Winterbottom, M. & Wilkins, A.J. (2009). Lighting and discomfort in the classroom. *Journal of Environmental Psychology*, 29, 63–75.

Received 23 February 2009; revised version received 3 March 2009.

Address for correspondence: Arnold Wilkins, Department of Psychology, University of Essex, Colchester CO4 3SQ, UK. E-mail: *Arnold@essex.ac.uk*