

Robert Waller

Comparing typefaces for airport signs

Keywords: legibility, wayfinding, signage, transport information, typographic research

This study combined three research methodologies to inform the choice of a typeface for signs at London's Heathrow Airport. The methodologies were legibility testing, qualitative consumer research, and expert review. The study showed that, contrary to a number of expert predictions, the serified typeface performed as well as the sans serif in legibility testing. Character width was a more significant factor in legibility, with condensed sans serif performing relatively poorly. The use of multiple methodologies led to a richer basis for decision-making: the qualitative research revealed clear genre expectations among airport users for sans serif signs; the expert reviewers raised a range of additional issues of genre, culture and context.

Introduction

Background and context

BAA, operators of Heathrow and other airports, are building a new terminal. Designed by Norman Foster, Terminal 5 has been a massive project, and over 20 million passengers are expected to pass through it each year.

Effective wayfinding is obviously critical to the success of airports, and BAA has a specialist team responsible for the complex task of maintaining effective wayfinding systems across their airports. They use sign standards drawn up in the mid 1990s based on work by Henrion, Ludlow and Schmidt. These standards will be familiar to anyone who has flown through a London airport. Black on yellow, they are unusual in their use of a serified typeface known as BAA Sign.

BAA Sign was created as a signage font that would match the visual identity in place at the time, in which the corporate font was Bembo. The corporate font has since changed to Frutiger, but the thousands of signs across BAA airports have not been changed, largely for cost reasons, but also in the absence of a compelling functional reason to change.

The opening of Terminal 5, an advanced high-tech airport building, was an occasion to reconsider this. Does the sign standard itself represent best practice in design for wayfinding? Would Frutiger, the corporate font, and one used extensively in signage, be a better choice? BAA commissioned research to find out.

How to research the choice of typeface?

What is the basis for choosing a typeface for any particular application? Availability, attractiveness, fashion, genre

associations, brand identity, and legibility all play a part in the choice, and the justification for the choice (not always the same thing). We used a combination of methods that addressed all of these factors to some degree.

The research brief was highly focused – it was not an open search for the most legible font for signing, nor was it open to us to develop a new font. Instead it was a straight comparison between two contenders: the bold serifed font in current use at BAA's UK airports (known as BAA Sign), and Frutiger, the font used in their current visual identity. So the initial filtering was done for us largely for brand identity reasons, although genre associations and legibility also implicitly played a part: both fonts have a history of use in airports, and both lay claim to legibility testing during their development.

As well as typeface choice, the research also looked at colour combinations, comparing the current black on yellow with alternatives that included white on black, white on grey and black on white. We used three research methods:

- Legibility testing, in which we measured the recognition speed resulting from words displayed in each font.
- Qualitative research, in which we asked individuals to judge the connotations and genre associations of the fonts, and express preferences for use in airport signage.
- An expert survey, in which we asked a panel of recognized experts to comment on the fonts, and on other aspects of BAA's sign standards.

We did this because we predicted that legibility research alone might not be conclusive. Historically, the experience has been that legibility differences between fonts of conventional design are often quite marginal, and our own judgement was that BAA Sign and Frutiger would be quite close.

We included the qualitative and expert research to give BAA a wider range of evidence or opinions on which

to base their decision. They commissioned this research because any decision to change fonts or colour could be costly, and needed justifying on functional grounds as well as those of branding or personal preference.

Legibility testing

Methodology

Legibility research has a long history (going back to the 1870s). A wide range of issues has been studied, including type size, line spacing, line length, type style, serifs and more. However, as Buckingham (1931) pointed out relatively early on, these factors interact in complex ways apparently unrecognised by many of the researchers. Indeed, in recent times a consensus has grown that the interaction of variables in type design is so complex that few generalizable findings can be found (see a longer review in Waller 1990). Instead, modern thinking is that legibility research is best conducted to solve specific problems and to test specific typefaces for known purposes, particularly where legibility is a critical functional issue. Recent examples are the development of fonts for people with visual impairments (Perera 2001) and for use in highway signs (Garvey, Zineddin and Pietrucha 2001). This research follows in that tradition, and is a highly focused study designed to solve one specific problem

Experience shows that it is quite hard to show significant and noticeable differences in legibility between normal (that is, not decorative or distorted) typefaces when displayed in good reading conditions. For research on typefaces for continuous reading, it is possible to create a reading task that is sustained enough to make differences visible, by giving people a text to read that takes several minutes, and timing them. But for typefaces designed for signage or other reading tasks involving short words or phrases, most research methodologies choose to degrade

the reading experience in order to bring out legibility differences. This is valid, as in reality legibility matters most when signs are viewed at a distance.

A common technique for assessing legibility is distance testing. In its simplest form, the viewer moves towards the test object, until they can correctly identify its content. The typeface that can be accurately seen from furthest away is the most legible. Variants on this technique include: bringing the stimulus progressively closer to the viewer on a rail (Nilsson 1991), revealing the stimulus for a fraction of a second using a device known as a tachistoscope (this is more often used for testing individual letters) and using filters to progressively blur the image (Schieber 1994)

Our computer display technique

For this study we developed a similar methodology, using a computer display on which words were gradually enlarged until the participant was able to read them.

From a blank screen, a word was gradually enlarged until the participant could read it. He or she pressed a mouse button, which stopped the word and caused the screen to go blank (this was to prevent them taking more time to read it). The participant then said the word out loud, and the researcher pressed a key to confirm it was correct.

At this point, the program recorded a set of data in a spreadsheet that included the typeface/colour combination; the word displayed, the time taken to recognise it, and whether the reading was correct.

Incorrect readings were excluded from the analysis. One subject's data was rejected because she made excessive errors, but for the most part they were rare.

In order to prevent differences due to learning effects, we gave participants 10 practice words to start with, and did not record the data.

We ran a pilot study with 6 colleagues to test appropriate distances from the monitor, font size and

lighting. A distance of 3 metres from the screen worked well, and was practical given the size of room available to us. We adjusted the speed of enlargement and the gaps between the exposures so that they were comfortable for the participants.

The word set

The words used for the testing were carefully chosen. In order to prevent easy guessing, we excluded the normal set of airport-related words from the study (for example, Arrivals, Departures, Gate, Transfer, etc). Instead, we picked a bank of 100 words from the Dale-Chall list of the 3000 most common English words. We looked for words of equal length (9 letters) that had at least two ascenders or descenders. Ascenders are the parts of letters such as 'h' or 'd' that rise above the x-height (the term typographers use for the middle part of a letter). Descenders are the parts in letters such as 'q' and 'p' that descend below the baseline. Because it is a widely held theory that a distinctive outline shape contributes to word recognition, we wanted to ensure these important legibility cues were reasonably consistent across the test words (see Larson, 2004, for an excellent review of word recognition theories).

No words were repeated for each participant. They were picked randomly, so each condition (that is, each combination of typeface and colour) was judged with a range of different words.

Ecological validity

In an ideal world, it might be argued that signs should be tested in situ, in real settings with real users. However, this would be impracticable for several reasons, including the high cost of mounting signs with multiple font variants in turn, and the difficulty in obtaining judgements in consistent conditions. It would also be easy to guess word content from a blurred shape, using contextual knowledge of typical airport words.

Although our judgements were obtained sitting in a quiet room, reading from a computer display without the stress of catching a flight, the important thing is that all typeface variants were compared under equal conditions. It is the comparison that is important to this study, not absolute measures. Having said that, computer monitors are backlit so they create a reasonable simulation of an internally illuminated sign, and we controlled the ambient lighting so that it was at the level recommended by BAA for airport environments.

Selecting typefaces to test

BAA Sign was originally drawn for use by BAA, and was designed to be associated with Bembo, which at the time was BAA's corporate typeface. So branding played a large part in its choice, although legibility research of some kind was undertaken (unfortunately, we could not track down any report of it, and the data has been lost).

Frutiger was chosen as the alternative, because it is BAA's current corporate typeface, and it is already used as a secondary typeface in signage. Moreover, Adrian Frutiger originally designed it for airport use (Charles de Gaulle, Paris). We included two weights in the study – Frutiger Bold and Frutiger Roman.

Legibility research uses statistical tests to ascertain that differences found are not due to chance. Sometimes the differences found, although statistically reliable, are very small, and might not affect users in a real world situation. So for designers looking for practical guidance, the question arises: is the difference enough to matter? To see if the results are of practical significance, we included two other typefaces in the study – one which we did not expect to be legible, and another with strong legibility credentials. In this way, we established a metric for the study to help us put any differences we found into perspective.

The low-expectation typeface was Stempel Garamond Italic. Although it is beautiful and elegant, no reputable designer would choose it for legible airport signing. So

for our methodology to be seen to work, and to claim credibility for practical decision-making, it had to show up a noticeable legibility deficit.

The high-expectation typeface was Vialog. This was designed for use for transport information, and was reportedly tested for legibility by its creators (Linotype 2005).

Colour combinations

Our brief was also to test the different typefaces in different colour combinations. These were:

- black on yellow (as currently used)
- black on white
- white on black
- white on grey.

BAA's visual identity specifies Pantone 123, which is a rich golden yellow. The nearest equivalent 3M film (commonly used in backlit signs) is Sunflower. We set the monitor to #FFC726 as the nearest to these. For the grey background we used the equivalent to Pantone 431, again a BAA standard grey. On the monitor this corresponds to #616A74.

When type is reversed (that is, displayed in white on a dark background), it tends to look bolder than the same font in black on white. We therefore compensated for this by creating special versions of the fonts with a 5% reduction in boldness.

The 20 conditions

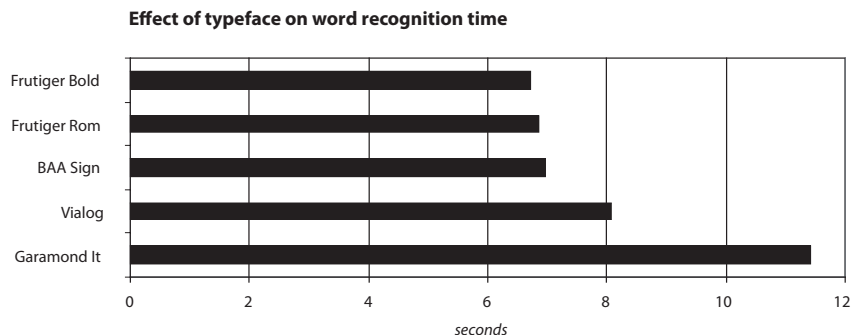
In summary, there were 20 conditions tested, with the Vialog and Garamond Italic ones only included to provide a metric against which to measure the others:

Participants

Twenty-three volunteers were recruited from among students at Luton University, 10 men and 13 women. They were given a sight test before taking part to ensure they had normal corrected vision. Twenty-five had originally

BAA Sign black on yellow	Frutiger Bold black on yellow	Frutiger Roman black on yellow	Vialog black on yellow	<i>Garamond Italic</i> black on yellow
BAA Sign white on black	Frutiger Bold white on black	Frutiger Roman white on black	Vialog white on black	<i>Garamond Italic</i> white on black
BAA Sign black on white	Frutiger Bold black on white	Frutiger Roman black on white	Vialog black on white	<i>Garamond Italic</i> black on white
BAA Sign white on grey	Frutiger Bold white on grey	Frutiger Roman white on grey	Vialog white on grey	<i>Garamond Italic</i> white on grey

Figure 1. The 20 conditions we tested



NB Throughout this paper, recognition time (in seconds) should be interpreted simply as an arbitrary comparative figure that is a function of this particular methodology.

Figure 2.

been recruited, but one failed to show, and another performed erratically and we rejected her data, leaving us with 23.

The experiment used a within-subjects repeated measures design, and this number of subjects was chosen to be sufficient to produce robust statistical data. Each of the 20 conditions was presented 5 times to each participant. This means we collected a total of 115 separate judgements about each from across the 23 participants. It also means that we collected 460 judgements about each typeface (ie, consolidating data from across the four colour combinations), and 575 judgements about each colour combination (ie, consolidating data from across the 5 typefaces).

Results: font comparison

Figure 2 shows the average recognition speed in seconds for each font tested. The shorter the line, the more legible the typeface.

An Anova statistical test confirms that the differences viewed as a group are statistically significant ($P=3.7446E-268$). That is, they are not the product of random factors.

We also used two-tailed t-tests to compare pairs of fonts. Figure 3 shows which paired comparisons were statistically significant.

This tells us that:

- Frutiger Bold is the most legible of the fonts tested, and is more legible than BAA Sign. However,

Overall results		Statistical significance of each paired comparison				
Fonts in order of legibility	Seconds	Frutiger Bold	Frutiger Rom	BAA Sign	Vialog	Garamond It
Frutiger Bold	6.707					
Frutiger Rom	6.854	×				
BAA Sign	6.980	✓	×			
Vialog	8.068	✓	✓	✓		
Garamond It	11.406	✓	✓	✓	✓	

✓ = highly significant ($p < 0.001$); × = not statistically significant ($p > 0.05$). 'Statistically significant' means that there is a very low probability that the results are due to chance.

Figure 3.

although the difference is statistically significant (that is, not the result of random variation), it is debatable whether it is great enough to be of practical significance in the airport environment.

- Frutiger Roman and BAA Sign must be regarded as equally legible – the small difference in the results could be the result of random factors.
- Vialog is definitely less legible than BAA Sign and both variants of Frutiger. In view of its development for transportation information, this is surprising, and we discuss possible explanations below.
- Garamond Italic was considerably slower to read than the other fonts. This result is only included to reinforce our confidence in the methodology.

Interestingly, these results were almost exactly mirrored in the pilot study where we used just six of our own staff to test the methodology.

Comparability of font sizes

Legibility research data has always been difficult to interpret and generalize from. One reason for this is the imprecision of typeface measurement.

Traditionally, font size (eg, 36 points) has described an invisible box that stretches from the top of the highest ascender (ie, the top of letters such as h, k, l, d, b) to the

bottom of the lowest descender (p, q). However, the part of the letter that dominates the apparent size, as judged by readers, is the x-height.

It follows from this that we have a choice of comparing fonts with the same nominal font size, or alternatively fonts adjusted so their x-height is equivalent. However, to do this is to make the assumption that the length of ascenders and descenders has no effect.

It is also probable that the width of typefaces has an effect on legibility, and this could explain the poor performance of Vialog in this study. BAA is designed to a generous width, and the results of the study show the widest typeface as the most legible, and the narrowest as the least legible.

Figure 4 shows BAA Sign followed by Frutiger Bold, Frutiger Roman, Vialog and Garamond Italic, superimposed on the outline of BAA Sign. It is clear that Frutiger has a slightly higher x-height, and BAA Sign is the widest of the fonts tested.

It should be pointed out that traditionally typefaces were redrawn for use in different sizes – in particular the length of ascenders and descenders, the junctions of curved and straight strokes, and serif shape and size would be adjusted to compensate for optical characteristics and the imprecisions of the printing process.

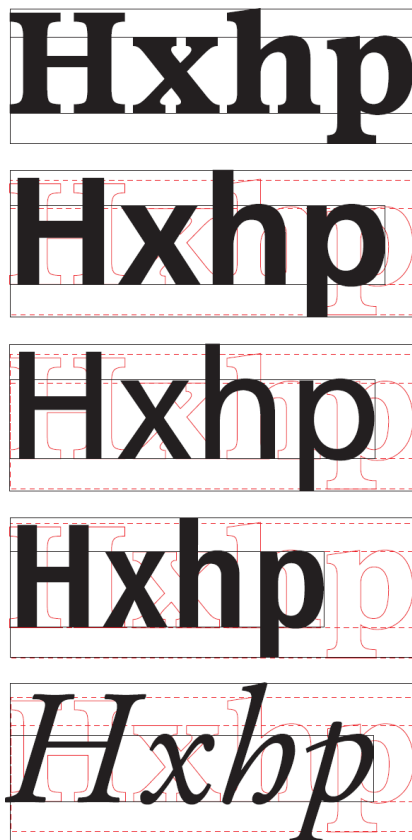


Figure 4.

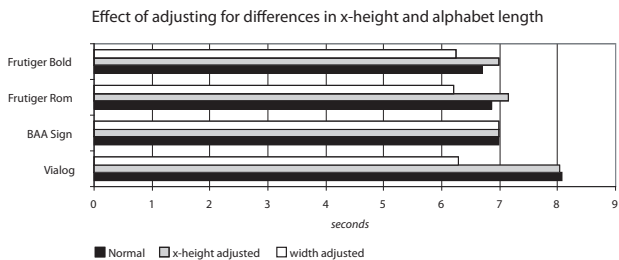


Figure 5.

Figure 5 shows the effect of adjusting the data we obtained to compensate for differences in x-height and width (as measured by differences in alphabet length). This is somewhat artificial, however, as it would not usually be realistic to, for example, enlarge Vialog so that its width is equivalent to that of BAA Sign.

Frutiger has a slightly larger x-height than BAA Sign – when we adjust the data to compensate for this, it removes the legibility difference altogether. Again, the reality is that we would not be able to fit an enlarged Frutiger into the same vertical space as BAA Sign, since the ascenders and descenders of adjacent lines would be too close. So the original font size comparison is of more practical relevance in this particular instance.

Implications for the use of condensed type

Condensed type is designed to save space, but this finding suggest that it is at the expense of legibility. In other words, a smaller size of non-condensed type could achieve the same result. It would be interesting to do further experiments to look at the effect of condensed type on continuous reading.

This does not mean condensed type has no use – in some contexts, it is important to retain the x-height, either to match the height of another typographic element, or to make the sure the word or sign is conspicuous.

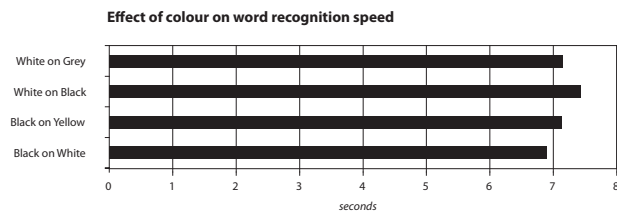


Figure 6.

Overall results		Statistical significance of each paired comparison			
Colourways in order of legibility	Seconds	Black on White	Black on Yellow	White on Grey	White on Black
Black on White	6.861				
Black on Yellow	7.137				
White on Grey	7.146				
White on Black	7.438				

= highly significant ($p < 0.001$); = not statistically significant ($p > 0.05$)

Figure 7.

Colour comparison

The average of all results split by font/background colour combination are shown in Figure 6.

Again, an Anova statistical test confirms that the differences viewed as a group are statistically significant ($P = 0.007807$), but at a lower level than the font comparison.

Figure 7 shows the results of paired comparison t-tests. These results reflect the findings generally found in the legibility research literature: that is, that contrast is the most important factor (black on white being the best), with negative formats (eg, white on black) being less legible than positive (eg, black on white), with the exception of people with certain types of visual impairment, for whom background glare is a problem (Silver, Gill & Wolffsohn, 1994, reported that people with macular disease, cataracts, and simple presbyopia strongly preferred light characters on a dark background).

We might have expected the white on grey condition to be less legible than white on black, since it is lower contrast. The explanation could be in the dazzle experienced with back-lit displays, which would be worse in the case of white on black.

The qualitative research

Because we predicted that the legibility results would be close, we planned additional qualitative research to

look at the personality communicated by typefaces, and specific preferences for airport use. There is a long history of research into the personality of typefaces (for example, Ovink 1938, Rowe 1982, Bartram 1982). We used two different approaches: one asked users to make judgements about abstract personality dimensions of typefaces, and the other asked them to rate their suitability for a range of specific functions.

Questionnaires were filled in on-line by around 400 participants, divided between the UK and Germany, reflecting the fact that Heathrow Airport users are as likely to be from outside the UK as from within. We tried to ensure a good mix of ages, and to include roughly equal numbers of men and women, and we included only people who had travelled abroad in the last year. They were recruited from a research panel organized by Lightspeed Research, a specialist company we used to collect the data.

We used foreign participants as well as UK ones, because BAA Sign is a familiar font for most UK travellers, and this fact could bias their view of its suitability for airport use. The choice of Germany as the non-UK location was on cost and convenience grounds. Lightspeed Research were able to offer Germany as well as the UK without having to subcontract to other companies. Although it might have been of interest to include people from a culture where the Roman alphabet does not predominate, that would extend the scope of the study into issues it was not designed to address.

The typographic personality question

We used bipolar scales to record judgements about the personality of each typeface. We based the scales on two sources:

- Past unpublished research, where we asked individuals to generate their own terminology for describing typefaces: Old-fashioned–Modern, Dull–Lively,

Informal–Formal, Technological–Human are typical examples of terms people readily use to describe type.

- Values which we judged would be relevant to BAA’s brand and reputation: Welcoming–Unwelcoming, Efficient–Inefficient, British–International, Straight-talking–Bureaucratic were chosen for this reason.

The full set was:

Old-fashioned	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Modern
Welcoming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Unwelcoming
Dull	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Lively
Straight-talking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Bureaucratic
Informal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Formal
British	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	International
Technological	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Human
Efficient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Inefficient

Neutral judgements

Past research indicates, as one might expect, that people find it harder to make judgements about typefaces with fewer obvious quirks or memorable characteristics. In this study, there were quite a high proportion of neutral judgements (ie, where the middle point in the scale was selected). This was particularly true of Vialog and BAA Sign, whereas people were most likely to form a stronger judgement about Garamond Italic (not surprisingly), and also Frutiger Bold (quite surprisingly).

We included Garamond Italic for interest, although it is not a practical choice for airport signs, and to give the participants a sense of difference. We have previously found that if you give people a set of similar sans serif typefaces, they sometimes have difficulty seeing any differences.

Frutiger Bold

To prevent confusion between the two weights of Frutiger, we only used Bold for this part of the study. It was seen as efficient, dull, formal, and British (which it

is not). British participants saw it as straight-talking and welcoming, while the Germans saw it as bureaucratic and technological.

BAA Sign

This was seen as old-fashioned, welcoming, fairly human while efficient, and formal. British respondents, but not the Germans, saw it as particularly British. The British saw it as somewhat dull and formal.

Vialog

This was seen as modern, efficient, and technological. The British thought it straight-talking. The Germans thought it dull, formal and less modern.

Garamond Italic

This was judge to be lively, human, welcoming, old-fashioned. The German respondents, perhaps surprisingly, saw it as informal and straight-talking. Interestingly, the British respondents were very polarized on the formal-informal dimension, with fewer opting for the ‘neither’ option.

The context of use question

As well as asking about abstract dimensions of personality, we also gave people a set of contexts in which they might expect to find the typeface. The question asked was “Where would you expect to find this font?” The range of contexts included a range of general uses, three of which were signage applications. At this point in the survey, participants were not aware of the context in which we are interested.

- packaging for bread
- logo for a restaurant
- headline in a magazine
- signs in an airport
- advert for a fast car

- logo for a computer company
- traffic signs
- signs in a shopping centre
- advert for a classical music concert

Frutiger Bold

Frutiger Bold scored highly on all three of the signage applications asked about. German respondents were more definite in this response than British ones. It also scored highly for magazine headline, but not for uses where a differentiated personality would be desirable (ie, the branding applications).

BAA Sign

The ‘very likely’ choice was not picked by many for any application, although it was considered ‘possible’ for most. However, it was picked as ‘fairly unlikely’ to be chosen for airport signs by quite a few respondents.

Vialog

Vialog was not chosen as ‘very likely’ by many (particularly in Germany), perhaps an indication that it was not much liked. It performed quite strongly on the signage applications but not as strongly as Frutiger Bold.

Garamond Italic

As expected, Garamond was not seen as a contender for signage, but was picked for some of the branding uses. There were a fair number of ‘possible’ votes for shopping centre signs (demonstrating that respondents were being quite discriminating in their choices).

Overall preference for airport signs

At this point we revealed that we were interested in airport signs, and asked for their preferences. For this question we introduced Frutiger Roman into the field. As Figure 8 shows, there was a clear preference for Frutiger Bold, although it was interesting that BAA Sign was the second choice, even among the German respondents for whom it was less likely to be associated with the airport context.

Preferred colour combination

We showed the set of colour combinations used for the legibility research, and there was an overwhelming preference for black on yellow. It is particularly striking that this very strong preference was among both nationalities, in spite of the fact that most German airports do not use

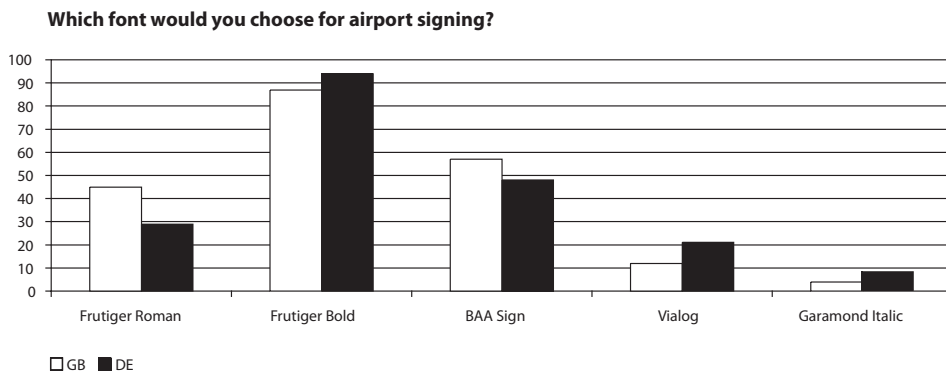


Figure 8.

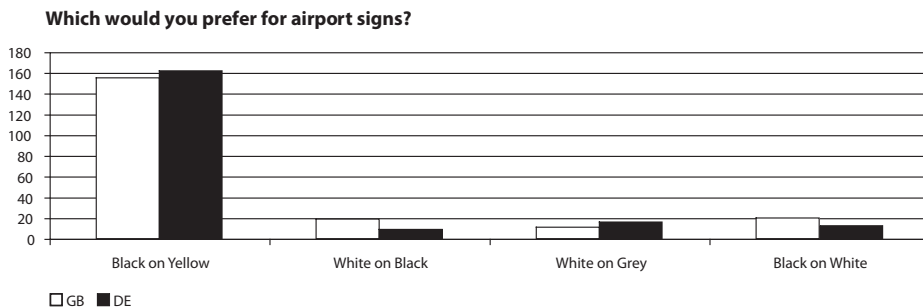


Figure 9.

black on yellow. In fact, most European and North American airports use white or yellow lettering out of black or dark blue. BAA and Schipol are the main exceptions.

This finding is so strong that we believe it needs further investigation – for example, presenting the stimulus in a real context to see the effect of surrounding visual noise.

The expert survey

The value of expert judgement is sometimes undervalued by professional researchers who are looking for theory-based, methodologically sound data. This issue was extensively debated in various multidisciplinary conferences held in the 1970s and 1980s (e.g., Kolers, Wrolstad & Bouma, 1979; Duffy & Waller, 1985). A strong advocate for the recognition of what he termed *master performers* was Michael Macdonald-Ross (Macdonald-Ross & Smith 1974; Macdonald-Ross & Waller, 1976), who argued for methodologies that harness the best of both sides. That is, solutions should be both well-argued, tested and methodologically sound, and also directed at practical issues of real relevance in the field, and making full use of the real but tacit knowledge of expert practitioners.

As the third prong in our methodology, then, we asked a number of experts in the fields of typeface design, wayfinding and legibility research to respond to a questionnaire.

The choice of BAA Sign for airport signs

The big issue is the personality of the font. It looks too old-fashioned, which may not be the right impression for a major international airport. This is a missed opportunity by the BAA to convey a clean, contemporary image, using a font that doesn't have so much personality but is visually simple and efficient, just as the airport experience should be. (Mark Ross)

The expert reviewers agreed that BAA Sign is not an obvious choice for airport signage, although two mention some advantages with the font, such as its generous x-height, weight, sturdy strokes and large counters.

However, it was also thought that its sturdy strokes and its relatively tight letter and line spacing make the font appear heavy and cluttered, and may reduce legibility at a distance. Others pointed out that the serifs add a certain amount of fussiness to the font. Erik Spiekermann expressed the view that even though serifs may help word recognition, signs are not read like continuous text where words are read by their outline shapes, but rather deciphered letter by letter (the evidence for this is not clear, but it does seem a reasonable model for non-English speakers trying to recognise words that they do not know how to say. Larson, 2004, reviews alternative word recognition theories).

Several of the expert reviewers argued that BAA sign is also an impractical choice. Because the font is voluminous and has a large footprint, it takes up more space

and is difficult to use when fitting long messages on signs. Other fonts would fit more characters in less space, and therefore be more flexible, effective and economical. This research casts doubt on that, since it shows that condensed fonts fit in more characters at the expense of legibility.

Nearly all had an opinion on the font's personality, and whether it is appropriate for airport signage. Most agreed that the font looks old-fashioned and seems to be a leftover from the 1960s and 1970s (although it was introduced in the early 1990s). They argued that the dark, complex and distinct font might be a good font for BAA branding literature, but that airport signs need to be simpler. Instead, the reviewers suggest using a sans serif font with an even stroke and medium weight. Barry Gray felt that a sans serif font would look clearer and more authoritative, which is important in this environment. Erik Spiekermann pointed out that nowadays, people are used to sans serif type on signs and may find a serif font strange and out of context.

However, it was also argued that there are advantages with BAA Sign. Because it is unique to the BAA, it gives the airport an identity, and distinguishes it from other airports around the world.

I like the different 'feel' of the signage – the Britishness of it is a plus. There are enough blandly signed airports around the world. (Gerard Unger)

Romedi Passini, on the other hand, argued that the font is not really a major issue in wayfinding, because it is more important to have the appropriate information at the right place for people to get to their destinations efficiently. In his view, the content and the location of the information are more important factors than the choice of font.

The legibility of BAA Sign compared to Frutiger Bold

[When equated for x-height] BAA Sign would probably work better [than Frutiger] as it has a bigger footprint.

In a real sign situation the better spacing and clearer letter style of Frutiger would, I would anticipate, give better results. However, other sans serif fonts or fonts with limited serifs may be even better. (Barry Gray)

A majority of the reviewers predicted that BAA Sign would be *less* legible than Frutiger Bold by about 10-20%, although three thought they would be about the same or that BAA Sign would be slightly better.

The majority view perhaps reflects the received wisdom in the design profession that sans serif is more legible for signs. Linda Reynolds remarked, "I can't help thinking that the Frutiger ought to be more legible", but she correctly predicted that BAA Sign would in fact be slightly *more* legible due to its extra width. She also noted the complications with comparing the legibility of two typefaces with different x-heights.

Two of the reviewers argued that Frutiger Bold is not the best sans serif font for airport signage. Frutiger is a legible font, but because it was not specifically designed for the use on signs it needs some adaptation.¹ Jean François Porchez also argued that Frutiger is not appealing enough to make the BAA identity unique.

Romedi Passini argued that any difference in legibility for the two typefaces would not be significant enough to affect wayfinding effectiveness.

The functional suitability of BAA's black on yellow colour scheme

I'm accustomed to it, and it seems OK, although the yellow is clearly a leftover from the 1960s. (Gerard Unger)

It works well, except that it is very generic, which is perhaps a boon for passengers who go to many airports, but not so good for BAA's identity. (Erik Spiekermann)

All reviewers agreed that the colour combination gives maximum legibility and contrast. It was also thought that the yellow gives good conspicuity against the walls and

ceilings of a typical airport and picks out the signs from the surrounding visual clutter.

However, once again, all reviewers agreed that the font's personality and its appropriateness for airport signage should be questioned. Mark Ross argued that the colour combination works hard to attract attention and although this is critical in this environment, it may also seem visually jarring and offensive.

Conclusions

Each of our three methodologies addressed different aspects of the decision facing our client, and each added a different perspective. The table below summarises the results of the comparison between BAA Sign and Frutiger Bold.

Our recommendation to BAA was that the legibility research did not in itself provide strong enough evidence to change font. So any decision to change font, on the

basis of the customer and expert views, would be branded, not functional.

We also recommended that they continue with black on yellow, since this performed almost as well as black on white and was endorsed by the experts and by customers.

Reflections on methodology

Hundreds of decisions are taken daily about the choice of font, and users are rarely consulted – it is usually impractical and expensive. Moreover, as the history of legibility research shows, it is often inconclusive.

On the other hand, while expert judgements were shown to contribute a richer set of cultural and genre factors to the decision, their views on legibility varied quite widely, and were unreliable – or perhaps judgements about suitability (which could be connected with genre or fashion) might here have been expressed in terms of predicted legibility.

	BAA Sign	Frutiger Bold
<i>Legibility</i>	Although statistically significant, the differences between these fonts were not big enough to matter.	
<i>User judgements about personality</i>	'Welcoming', and 'human' while 'efficient' and 'formal'. But the high score for 'old-fashioned' is a worry.	'Efficient', 'straight-talking', 'formal' – basically functional.
<i>User judgements about context of use</i>	Picked as a 'fairly unlikely' choice for airport signs, and was a lukewarm 'possibly' choice; for other signage applications.	Scored highly for all signage uses.
<i>User preference for airport signs</i>	Second choice.	First choice, with Frutiger Roman as third choice.
<i>Expert judgement</i>	Not the experts' choice, although an acceptance by some that it at least differentiates BAA airports. But they criticize its legibility (unfairly as it turns out) and articulate the users' judgement that it is dated.	Of the two, the experts' choice, although they would like to see adjustments to it – and perhaps there was unarticulated hope that BAA might commission its own font, as other airport operators have done.

The study produced a simple legibility testing method, which should be relatively easy to replicate, although it should be noted that it was specifically designed to replicate the way signs are read. It shows how single words are read (and it could be used with short phrases, too) and the results should not be transferred across to small text for continuous reading on paper or screen.

Some of these findings could be followed up by further research. In particular, our finding on the legibility of condensed type should be followed up, as every guideline on legibility (including guidelines for large print for visually impaired people) focuses strongly on type height alone. The strong preference for black on yellow signs might also be followed up, to see the effect of context.

The attempts of past researchers to discover universal rules for typography have rightly been consigned to history, but targeted legibility research does have a place in the practice of design. Simple, pragmatic methodologies, like the one we developed for this study, are particularly useful where experts cannot agree, where clients need reassurance, or where legibility is particularly critical to successful communication.

Acknowledgements

Thanks to Mike Wolff of BAA for commissioning this research and giving permission to publish it. Thanks also to colleagues Alison Dixon, Hanna Fritzell and Robert Laing; to Jo Allerton of JA Ergonomics who ran the study and analysed the data; to Lightspeed Research who conducted the web survey; to our panel of experts who kindly agreed to offer their comments and have them published. They were: Barry Gray, Chair, Sign Design Society, UK; Professor Romedi Passini, Université de Montréal, Canada; Jean François Porchez, Porchez Typofonderie, France; Linda Reynolds, University of Reading, UK; Mark Ross, Dot Dash Design, Brisbane, Australia; Professor Erik Spiekermann, United Designers Network, Germany; Professor Gerard Unger, University of Reading, UK.

Note

1. In fact, it was specifically designed for airport use – for Charles de Gaulle airport in Paris. However, what this comment is probably referring to is the fact that the particular cut of Frutiger we tested was not designed for signs.

References

- Bartram, D. (1982). The perception of semantic quality in type: differences between designers and non-designers. *Information Design Journal*, 3, 38–50.
- Buckingham, B. R. (1931). New data on the typography of textbooks'. *Yearbook of the National Society for the Study of Education*, 30, 93–125.
- Duffy, T., & Waller, R. H. W. (Eds.) (1985). *Designing usable texts*, Orlando: Academic Press.
- Garvey, P. M., Zineddin, A. Z., & Pietrucha, M. T. (2001). Letter legibility for signs and other large format applications. *Proceedings of The Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1443–1447).
- Jonassen, D. (1982). *The technology of text*. Englewood Cliffs: Educational Technology Publications.
- Kolers, P. A., Wrolstad, M. E., & Bouma, H. (1977). *Processing of visible language*, 1. New York: Plenum Press.
- Larson, K. (2004). *The science of word recognition*. Advanced Reading Technology, Microsoft Corporation. (www.microsoft.com/typography/ct-fonts/WordRecognition.aspx)
- Linotype (2005). www.linotype.com/53981/vialog-family.html
- Macdonald-Ross, M., & Smith, E. (1974). *Graphics in text: a bibliography*. The Open University Institute of Educational Technology.
- Macdonald-Ross, M., & Waller, R. H. W. (1975). Criticism, alternatives and tests: a framework for improving typography. *Programmed learning & educational technology*, 12, 75–83.
- Nilsson, T. H. (1991). *Legibility of tobacco health messages with respect to distance*. Report to the Tobacco Products Division of Health & Welfare Canada (contract H4078).
- Ovink, G. W. (1938). *Legibility, atmosphere-value, and forms of printing types*. Leiden: Sijthoff.
- Perera, S. (2001). *An investigation into the legibility of large print typefaces*. UK: RNIB Scientific Research Unit.

- Rowe, C. L. (1982). The connotative dimensions of selected display typefaces. *Information design journal*, 3, 30–37.
- Schieber, F. (1994). Using the “blur tolerance” technique to predict and optimize the legibility distance of symbol highway signs. *Proceedings of the Human Factors Society*, 2, 912–915.
- Silver, J. H., Gill, J. M., & Wolffsohn, J. S. W. (1994). Text display preferences on self-service terminals by visually disabled people. UK: RNIB Scientific Research Unit
- Waller, R. H. W. (1990). Typography and discourse. In R. Barr et al. (Eds.), *Handbook of reading research* (pp. 341–380). New York: Longman.

About the author

Robert Waller is Head of Information Design at Enterprise IG. He began his career as a researcher with the Open University, before starting an information design consultancy, Information Design Unit (acquired in 2001 by Enterprise IG), in 1988. He was the founder and first editor of *Information Design Journal* in 1978, and has a doctorate from Reading University for a study of the relationship of typography and language.

Contact

Robert Waller
Head of Information Design
Enterprise IG
11-33 St. John Street
London EC1M4PJ
United Kingdom
www.enterpriseig.co.uk
e-mail: rob.waller@enterpriseig.com