Simplification
centre

Benchmarking everyday documents

Martin Evans   *April 2011*

This paper explains how the Simplification Centre benchmarks documents for clarity and usability. We rate documents on sixteen research-based criteria and weight the results to give an overall score.

Results from our first eight reports gave overall scores between 35% and 70%. We found few problems with grammar and punctuation, but more documents failed to make it clear what the user needed to do with them. The biggest challenge was to find a clear organising principle and use design and layout to guide the user through.

University of
Reading
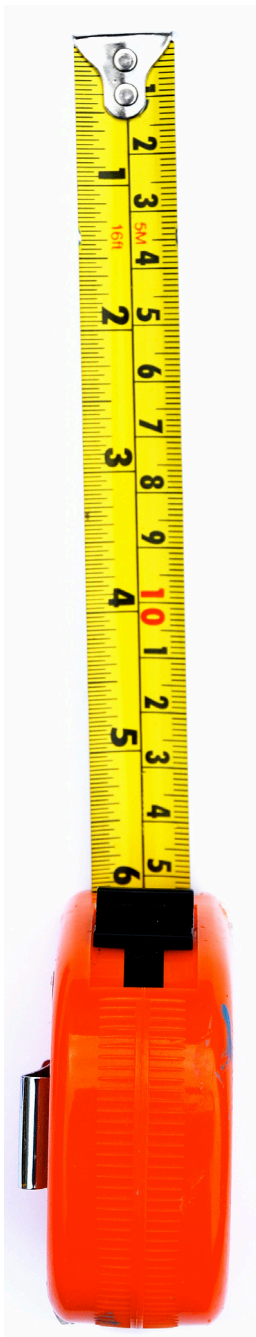
Why and how we benchmark

Many of the everyday documents people have to deal with fall far short of the best practice in clarity and usability. Yet there exists a considerable body of knowledge, based on research, about what makes documents easy or hard to understand. Part of the Simplification Centre's purpose is to bridge that gap – to help people who produce documents to apply existing knowledge in order to raise the general standard of usability.

Why we benchmark

One tool to help with this is benchmarking. It helps organisations to know how their products compare – with others and with current best practice. Benchmarking is most effective when it forms part of an organisation's process of continuous improvement. So the Simplification Centre set out to offer its member organisations[1] a document benchmarking service, and to make it as useful as we can to our members' own efforts to achieve high standards in their documents.

Our benchmarking reports consist of a mixture of scores, analysis and recommendations. Documents, or in some cases sets of related documents, are given a score for each of sixteen criteria. The assessors' scores represent their judgement in relation to the stated criteria and the descriptors for the scoring levels. This approach carries risks of subjectivity and inconsistency. We mitigate these by using a second evaluator to moderate the scores; by evidencing the judgements with specific examples; and by monitoring for consistency between reports. Nevertheless we do not claim that the scores are anything more than (well-informed) judgements.

Our benchmarking does not include testing the documents with users. It is always possible that however well a document follows theoretical best practice, users have problems with it. Conversely a theoretically less than perfect document may work well enough in practice. We always advise our members that our benchmarking is not a substitute for user testing.

....

1 During the first two years of operation (2009-10) we ran a membership scheme, under which member organisations funded our work, and received reports on the quality of their documents.

What does the benchmarking research say?

There is a good deal of research literature on benchmarking. David Longbottom of the University of Derby (2000) has done a literature survey which reviewed 460 papers by practitioners and academics. These divided broadly into case studies and theoretical developments. The theoretical analysis largely builds on the original work by Camp (1989) which defined benchmarking as an integral part of business process improvement. Most studies have focused on manufacturing businesses who have analysed their processes and compared them with comparators in similar or different businesses, in the search for greater efficiency.

Another strand of studies has identified a different approach. The proponents of business process re-engineering have argued that comparing yourself with competitors is an insufficiently radical competitive strategy. Their interest in benchmarking has therefore been more focused on metrics: how do the numbers compare? Although the Simplification Centre's benchmarking is essentially qualitative – the numbers are simply our scoring system – we fit more easily into this second strand.

Some criteria were relatively easy to comply with. Most documents made it clear who they were from and what they were about; and how to get in touch. Grammar and punctuation were generally not a problem; but there was much more difficulty in creating an engaging impression for the user, minimising unfamiliar technical language, navigating the user clearly through the material; and, hardest of all, creating a clear, consistent and logical overall structure.

Our benchmarking process

The sixteen criteria

We benchmark documents against sixteen criteria, which fall into four broad categories. Technical Paper 2 in this series sets out the reasoning behind each criterion.

Language

These criteria assess the various aspects of the use of language in the document.

Directness: using direct language to make clear who's doing what.

Plain words: extent to which the vocabulary is easily understood.

Grammar and punctuation: conformity with the practice of good standard English.

Readability: ease with which the reader can follow the argument of the text.

Design

These criteria assess the visual impact of the document and the way its design influences its usability.

Legibility: use of legible fonts and text layout.

Graphic elements: use of tables, bullet lists, graphs, charts, diagrams, illustrations, etc.

Structure: quality of the document's organisation in relation to its function.

Impression: attractiveness and approachability of the document's overall appearance.

Relationship

These criteria assess how far the document establishes a relationship with the users which supports them in taking appropriate action.

Who from: is it clear who is communicating?

Contact: whether there are clear contact points and means of contact.

Audience fit: appropriateness to the knowledge and skills of the users.

Tone: matching the style and language to the context.

Content

These criteria assess how far the content and the way it is organised deliver the document's purpose.

Relevance: how relevant the content is to the recipient.

Subject: whether it is clear what the communication is about.

Action: clarity about what action is required of the user.

Alignment: compliance with the organisation's intended aims and values.

How we score

We score documents against the 16 criteria, using a four point scale:

★★★★   **Excellent:** represents current best practice. Makes its full contribution to maximising the clarity and simplicity of the document.

★★★   **Good:** generally represents good practice and contributes to a clear and simple document but with exceptions. Capable of further improvement.

★★   **Fair:** of average quality. May have some good features but capable of significant further improvement. Not representative of current good practice.

★   **Poor:** unclear, confusing.

How we weight the scores

We then weight the scores on a four point scale:

Legibility   Structure   Action
■■■■ = x 4

These get the highest weighting. If you can't read the words, you can't use the document. If you don't know what to do with it, it's failed its purpose. And the more complex the document, the more vital it will be that there is a clear structure that the user can find their way through.

Plain words   Readability   Audience fit   Relevance   Alignment
■■■ = x 3

Also of major importance: vocabulary needs to be understandable, preferably without need for a dictionary or glossary. Content needs to be suitable for the audience, and it must be possible to follow the argument. And even if a document is clear, if it leads the user to conclusions that are out of line with the organisation's intention, it will have failed its purpose.

Directness   Impression   Contact   Tone   Subject
■■ = x 2

These criteria are less likely to completely defeat the users, but can cause them a great deal of unnecessary work or irritation. Examples: searching for contact details; an inappropriate tone of voice in a sensitive letter. But a forbidding impression may put the user off engaging with the document, or persevering to the end.

Grammar + punctuation   Graphic elements   Who from
■ = x 1

Finally, these criteria rarely cause users to fail to understand, but they do annoy. Bad grammar or non-standard punctuation, for example, rebounds on the reputation of the organisation, even if it does not obscure the meaning.

Constructing an overall score

We multiply scores by weightings to give a weighted score for each criterion.

We show the weighted score compared with the maximum weighted score for the criterion.

And we show the number of points lost, ie the maximum minus the actual.

For example:

	Score	Weight	Weighted score	Points lost
Criterion 1	★★	■■	4/8	–4
Criterion 2	★★★★	■	4/4	0

Finally we add the total of weighted points and convert to an overall percentage score.
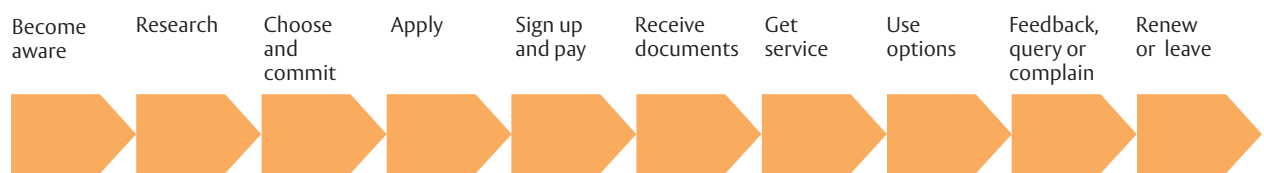
0% would be equivalent to a ★ score on every criterion.

100% would be equivalent to a ★★★★ score on every criterion.

The documents
and how they scored

The customer journey

Customers' experiences of using documents form part of their wider interaction with the organisation. This end-to-end experience is often described as the 'customer journey'. For the purpose of benchmarking, the Simplification Centre developed its own model of the generic customer journey. This is the current version:

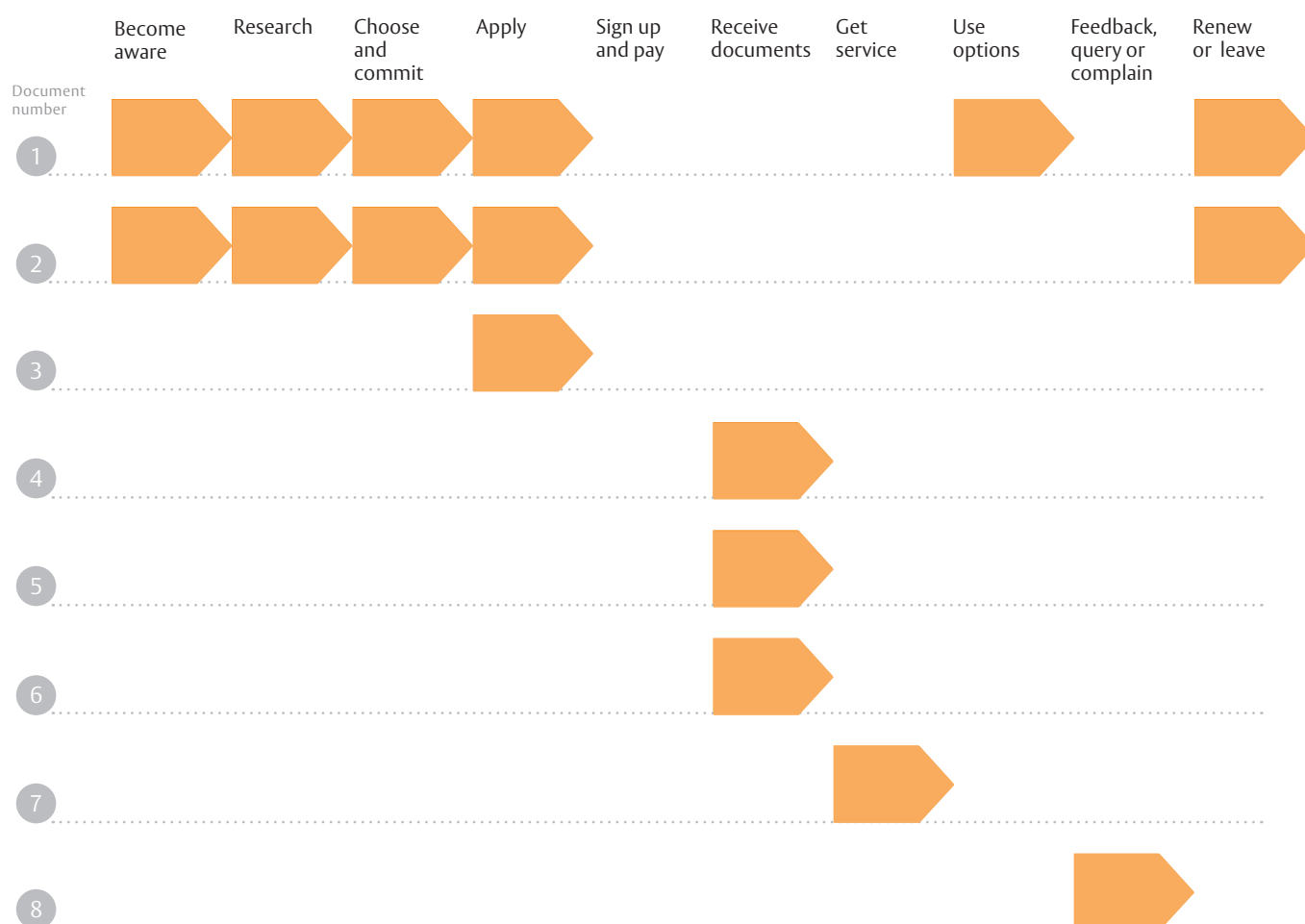| Become aware | Research | Choose and commit | Apply | Sign up and pay | Receive documents | Get service | Use options | Feedback, query or complain | Renew or leave |

Although the stages look typical of a customer purchase – buying insurance for example – they also work in less obvious contexts: 'Becoming aware', for example, would apply to reminders of the date that tax returns are due. You might undertake 'Research' to learn what VAT information you need to keep. A formal appeal against a government department decision is a form of 'Feedback query or complain' as is a customer satisfaction survey.

We take account of this context in our benchmarking. Documents have to be consistent with previous communications; and they may have a continuing role in informing later customer interactions. The more stages a document covers, the more it needs a clear structure, a strong narrative, and effective navigation.

The documents

We have mapped each of our eight benchmarking reports on to the customer journey model. Some of the documents or sets of documents we have looked at cover more than one of the stages. We aim over time to benchmark a reasonable range of documents across all the stages. This chart shows the current picture at July 2010. To preserve anonymity we have numbered the documents ① to ⑧.

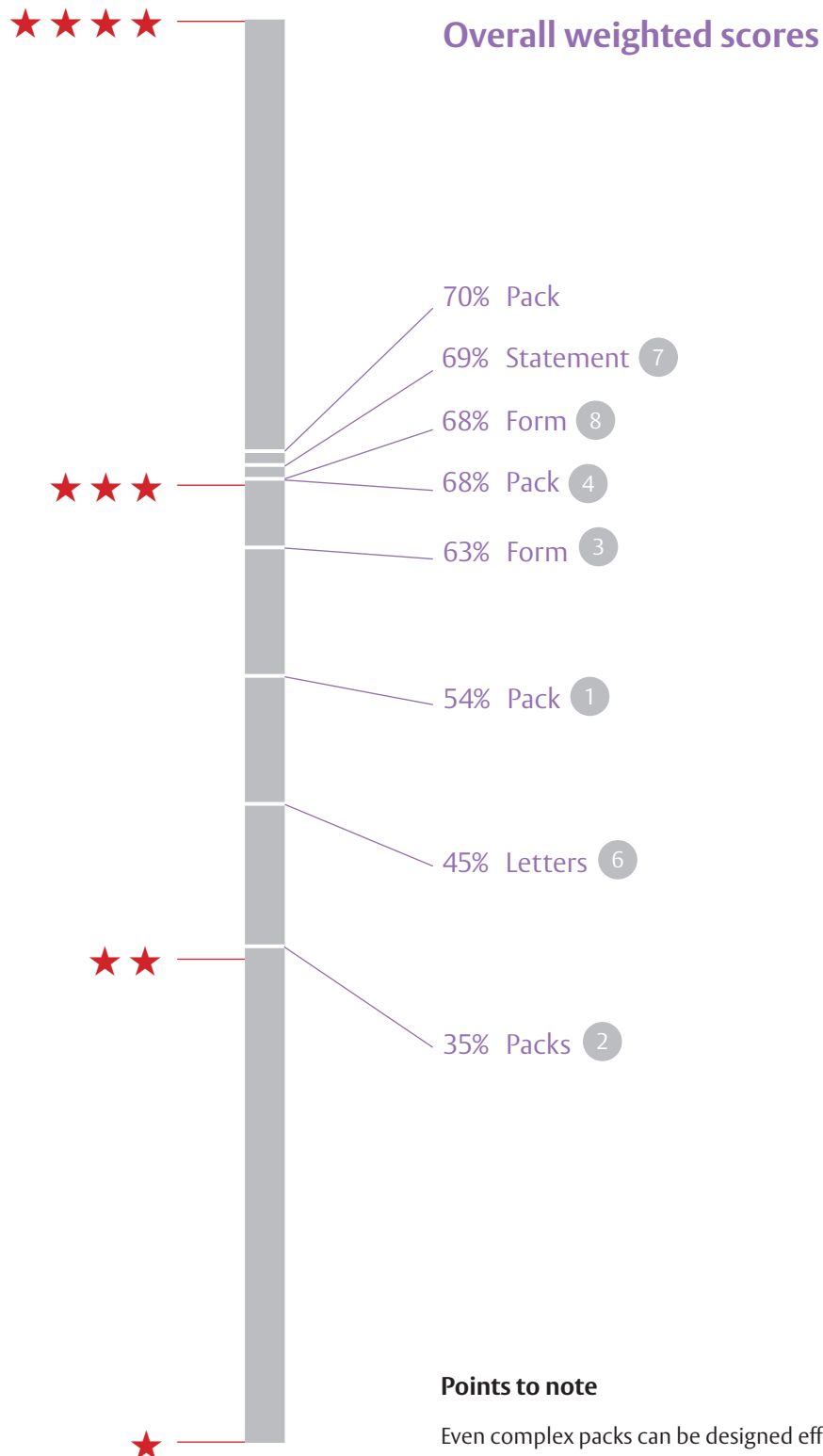| Document number	Become aware	Research	Choose and commit	Apply	Sign up and pay	Receive documents	Get service	Use options	Feedback, query or complain	Renew or leave
1	▶	▶	▶	▶				▶		▶
2	▶	▶	▶	▶						▶
3				▶						
4						▶				
5						▶				
6						▶				
7							▶			
8									▶	

Points to note

Two of the large document packs covered a whole series of stages in the customer journey. They were the type of document that would be referred back to for future reference, as well as having an immediate function. This is a major challenge and neither of these documents fell in the top range of scores.

The results

The 8 benchmark reports comprised 4 document packs, 2 forms, 1 statement and 1 set of letters. The overall weighted percentage scores of each are illustrated in this chart:

★ ★ ★ ★

Overall weighted scores

70% Pack

69% Statement ⑦

68% Form ⑧

★ ★ ★

68% Pack ④

63% Form ③

54% Pack ①

45% Letters ⑥

★ ★

35% Packs ②

★

Points to note

Even complex packs can be designed effectively.

All the documents tested were capable of further improvement in some of the criteria.

Adding up the scores

Some of the criteria proved harder to satisfy than others. The table below adds together the eight documents' unweighted scores for each individual criterion. On our four point scale, if all 8 documents scored the maximum ★★★★ then the total score for that criterion would be 24. Actual scores range from 11 to 20.

Criterion	Score
Subject	20/24
Who from	20
Grammar + punctuation	20
Relevance	18
Plain words	16
Audience fit	15
Contact	14
Legibility	14
Directness	14
Action	13
Tone	13
Structure	12
Readability	12
Alignment	11
Impression	11
Graphic elements	11/24
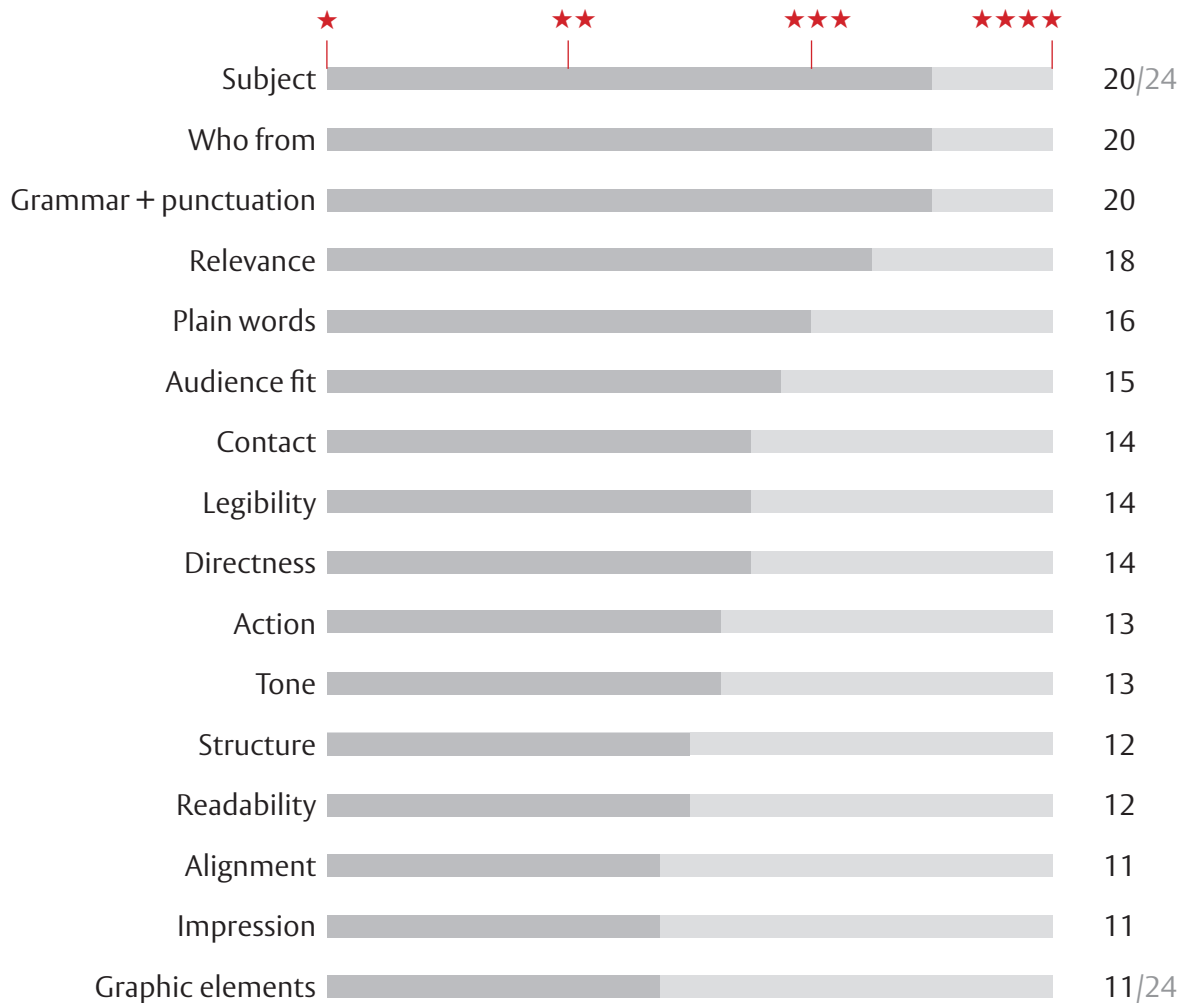
Point to note

On average, the documents handled clear language issues better than structure and visual design issues.

| Score	Value
★★★★	3
★★★	2
★★	1
★	0

The pattern of the scores

This chart

The chart shows, for each document, how many of the criteria achieved one, two, three or four star ratings.

The documents are shown in order of their overall weighted scores – from 70% down to 45%.

The document numbers ① to ⑧ identify the individual documents. These numbers are also used in the charts on pages 11 and 12.

The scores illustrated in the columns are the unweighted scores. So document ③ (overall weighted score 63%) appears to perform better than document ⑧ (overall weighted score 68%): this is because it scored more of its stars on low-weighted criteria and fewer stars on higher weighted ones.
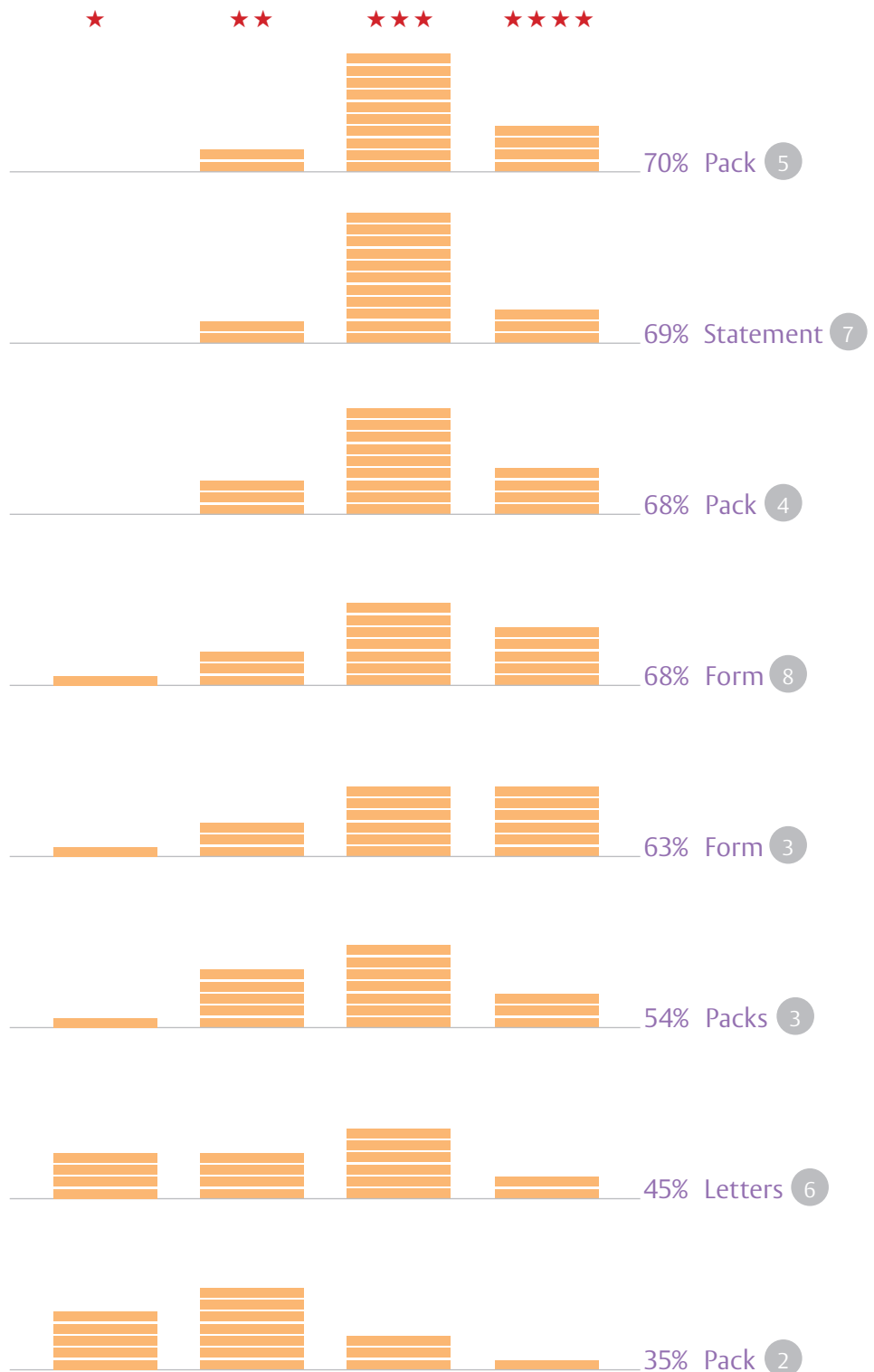
Points to note

The scores of documents and packs that had been the subject of efforts at improvement generally scored respectably or well. But even the best failed to reach ★ ★ ★ ★ level across the board. There was scope for further improvement in most criteria.

That is not to say that improvement is easy to achieve. In some cases the sheer complexity of the information to be communicated makes it extremely challenging to achieve simplicity and clarity.

★   ★ ★   ★ ★ ★   ★ ★ ★ ★

70% Pack ⑤

69% Statement ⑦

68% Pack ④

68% Form ⑧

63% Form ③

54% Packs ③

45% Letters ⑥

35% Pack ②

The criteria
How the documents performed

Language criteria

Directness: All the documents had made attempts to use direct language but in almost all cases there was scope to go further.

Plain words: documents generally used clear vocabulary appropriate to their users, providing definitions where needed. The challenge is to do so consistently. Almost always there is a lapse into technical or legal language at some points.

Grammar and punctuation: the quality of the unweighted scores is high in almost all cases. The weighting we give to this criterion is low because poor grammar or punctuation rarely impact on the usability of the document. But they do impact significantly on the reputation of the organisation.

Readability: the majority of documents fell below par here, at least in parts. Scores were low because of the difficulty of following the argument from sentence to sentence. The problem is one of structuring at the detailed level, to keep a narrative flow that can be followed without excessive cognitive load.

Design criteria

Legibility: Nearly all documents scored well but not excellently on legibility. Choice of font and type sizes were rarely a problem. The issues were more ones of typographic layout and particularly the use of space to make the structure visible.

Graphic elements: there was a general lack of well designed helpful tables and diagrams.

Structure: few documents scored really well on structure. The more complex the information, the tougher the challenge. There is a need to step back from the detail, see the big picture, and impose a clear organising principle, reflected in a small number of main sections: and then to organise each section appropriately.

Impression: a mixed picture here. We were not necessarily looking for glossy or striking graphics. But we were concerned that users' first impressions should not prevent them from getting engaged

with the content. Keeping bulk to the minimum helps. But there is a trade off with the need to create a visible structure, use space to express it, and avoid slabs of dense text. Some documents went to the other extreme of splitting text into a multiplicity of very short units, making it equally hard to see the wood for the trees. Use of colour helped where used consistently and with restraint: but hindered where it added to the visual noise.

Relationship criteria

Who from: this was almost always clear...

Contact: ...but it was not always clear how to get hold of them when you needed to. The best documents repeated the contact details at the points where users might need them.

Audience fit: the documents mostly did a respectable job of pitching their content at a level appropriate to their audience. The weaknesses were more at the detailed level: inconsistencies and lapses into technical or legal language at particular points in the story. Tackling this is an organisational challenge requiring team working.

Tone: there were few major problems of overall tone. The problems, again, were more of consistency. Documents tend to change voice from part to part. This is hard to avoid unless there is a strong unifying editorial hand

Content criteria

Relevance: scores here were generally good, with some particularly good examples of using personalisation of individual documents, to avoid loading unnecessary complexity on people who do not need it.

Subject: it was almost always clear, or very clear, what the document was about.
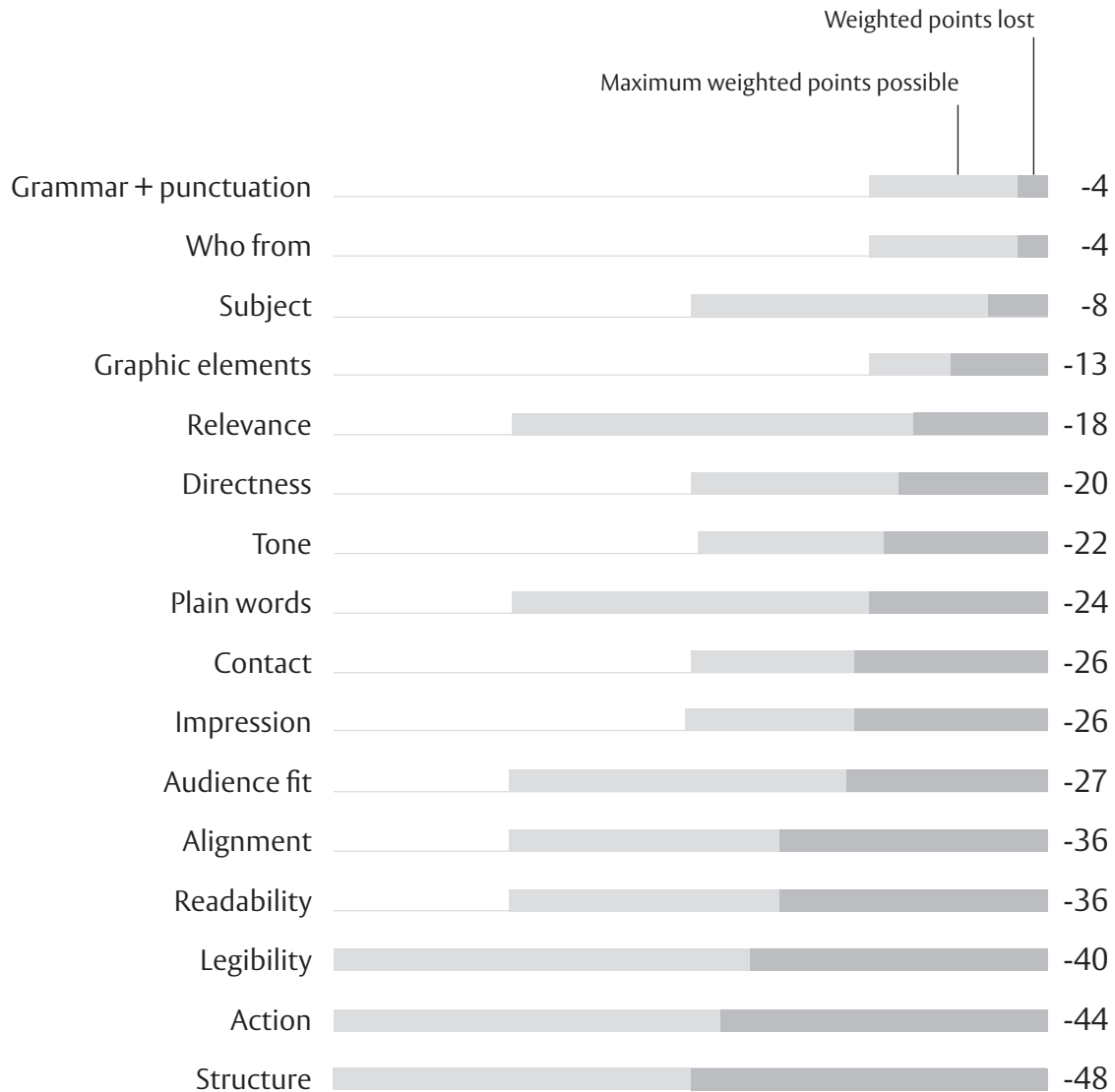
Action: but it was often less clear exactly what the recipient needs to do about it – and when. The best examples led with a short clear statement up front, setting out the action needed. This was supplemented by a checklist of actions linked to the final action.

Alignment: this was sometimes difficult for us to judge. A document may be clear to the user but have effects that are not in line with what the organisation is trying to achieve. In judging this we depend on what we are told about the aims. There is a general need for organisations to be conscious of how far their functional documents are actually aligned with their stated aims and values.

Weighted scores
Points lost = difficulties for users

Weighting the scores gives the best impression of the relative impact on the usability of the document. The table below shows the aggregate, over the eight documents, of the weighted points lost. Points lost are the difference between the points scored and the maximum possible points score for that criterion.

Weighted points lost

Maximum weighted points possible

Grammar + punctuation	-4
Who from	-4
Subject	-8
Graphic elements	-13
Relevance	-18
Directness	-20
Tone	-22
Plain words	-24
Contact	-26
Impression	-26
Audience fit	-27
Alignment	-36
Readability	-36
Legibility	-40
Action	-44
Structure	-48

Conclusions

The priorities for action

The Simplification Centre's benchmark process has enabled us to produce useful reports for members, feeding into their processes of document redesign. We do not have enough data yet to draw firm conclusions from the findings, but already there are some emerging patterns.

The table of weighted scores suggests that the three criteria where improvements would make the biggest impact are Structure, Action and Legibility.

Of these **Action** may be the easiest to tackle. The best documents resisted the temptation to begin with background explanation, expressions of commitment to quality, or other matters. They stated prominently and plainly why they were communicating and what the recipient needed to do about it. There was clear navigation to the action points in the document, and where necessary a checklist.

Legibility: the challenges here are of typographic design and layout. The strictly functional documents from financial services and government organisations rarely have much input from the graphic designers who deal with marketing or campaign publications. There is a widespread need for upskilling relevant staff, and a number of our member organisations are actively engaged in this.

Structure is probably the hardest challenge, being a skill of its own. Improving structure means challenging the form of documents that may have just grown over time. As well as skilled individuals who can see the wood for the trees, creating a complex document requires senior ownership in the organisation, a strong editorial function, and a willingness of the various stakeholders to work as a team.

References

Longbottom, D. (2000) Benchmarking in the UK: an empirical study of practitioners and academics. Benchmarking: an International Journal, 7 (2), 98-117.

Camp, R. (1989) Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance, ASQC Quality Press.