

Retrieved 28.6.2010 from: Communication Research Institute website
<http://communication.org.au/publications/communication-benchmarks/credit-card-statements--communication-benchmarks-2009/120,67.html>.



Credit Card Statements: Communication Benchmarks 2009

By David Sless and Alex Tyers

This is a summary of the 'headline' data collected in this study. If you would like access to the full data and our analysis of what needs to be done to improve specific credit card statements, please contact [David Sless, our Director](#).

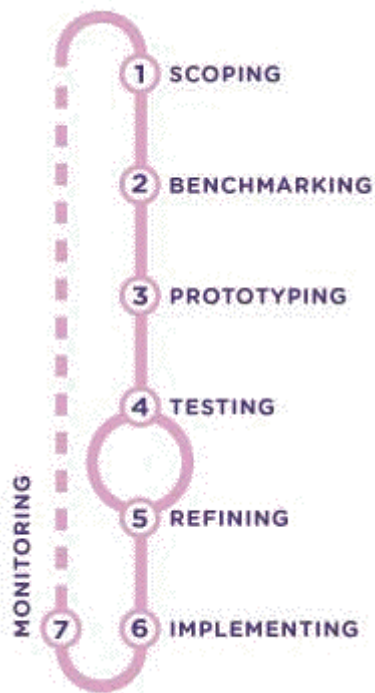
If you are interested in participating in our Communication Benchmarks Projects as a volunteer, or in any other capacity, please contact [Alex Tyers, our Communications Benchmarks Manager](#)

What are communication benchmarks?

Over 20 years ago the Communication Research Institute (CRI) began Communication Benchmark studies to measure the quality of communication practices used by business and government in their communication with the public.

The types of communication studied included such things as voice systems, forms, legal documents, bills, letters, product labelling, consumer instructions, and websites the stuff of ordinary life that originates from business and government and makes up a large part of the daily communication between organisations and the public.

Figure 1: The information design process



Most of these benchmarking studies have taken place in the context of CRI's ongoing program, helping its Member organisations improve their communication with the public.

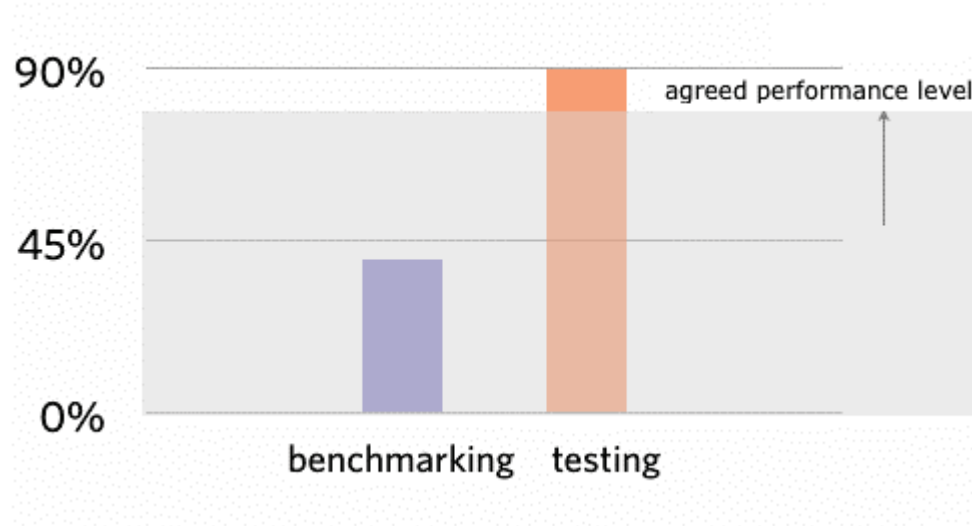
Benchmarking normally occurs at a particular stage of a communication or information design process when the current faults in a communication are identified through diagnostic sessions, before any new communication or information design work is undertaken.

Benchmarking sessions quantify the number and types of faults in a design and how far short they fall of an acceptable performance level. They also provide a great deal of qualitative data on the causes of the failure.

When used as part of a design process (Figure 1), the diagnostic data provided by such sessions enable information designers to move rapidly to rectify the faults in any redesign process. Additionally, data collected at the benchmarking stage can be used by information designers at the Testing and

Monitoring stages of the process to detect and quantify any improvements resulting from the new prototype (Figure 2).

Figure 2: Measuring outcomes at different stages



Background to this study

Over the last 10 years CRI has expanded the scope and application of its benchmarking studies, taking them out of their specific role within the design process for particular information documents, and using them as stand-alone studies of whole classes of information, calling these studies Communication Benchmarks.

In the mid-1990s CRI undertook Communication Benchmarks studies in Australia of banking websites (1), medicine labelling (2), Financial Services Guides (FSG) for financial products (3), government and business forms (4), utility bills (5), Consumer Medicines Information (CMI) (6), and many other types of designed information. Detailed data from these studies are available to [CRI Members](#). Wherever possible, we avoid drawing attention to specific institutions. We have no interest in 'naming and shaming'. Rather, our interest is in drawing attention to current public communication practices, in order to help the whole of industry improve its practices in the future.

The data we provide establish the communication benchmarks against which we can measure future improved practices.

This Communication Benchmarks is the first in which CRI has conducted such a study internationally. We chose for this first international study a document that is relatively widespread around the world, and which has attracted considerable public attention recently Credit Card Statements (CCS)

All those involved in the benchmarking activities were volunteers. Eleven CRI Fellows and Subscribers from Australia, Austria, Chile, Netherlands, Portugal, South Africa, UK and USA gave freely of their time as investigators on the project, and they recruited other volunteers to participate in the study. The whole project was managed by Alex Tyers in Melbourne.

Figure 3: The investigators

- Alexander Tyers, Australia
- Carola Zurob, Chile
- Claudine Jaenichen, USA
- Consuelo Amenabar, Chile
- Frances Gordon, South Africa
- Judith Moldenhauer, USA
- Karel van derWaarde, Belgium
- Martin Gallo, Argentina
- Sandra Fisher-Martins, Portugal
- Thomas Bohm, UK
- Veronika Egger, Austria

This is a report of their work; it has been made possible as a result of their collective efforts and collaboration. Both the investigators and participants have made an important contribution to our field, and we hope they will continue to collaborate with us in the future. Indeed, we hope that their example will encourage many others to join us in our Communication Benchmarks program.

Method

Diagnostic testing sessions

The method used in this study is called diagnostic testing (7). It is conducted in a session involving an investigator and a participant. As its name suggests, the method was developed specifically to help information designers identify and diagnose design faults.

There is now a body of evidence and experience in the use of diagnostic testing that confirms its technical reliability, sensitivity and validity (8). There are also good research conclusions on the best types and numbers of people that are needed as participants, in order to get useful data (9).

Like diagnostic tools used in medicine, diagnostic testing in information design is at its most powerful when used in a clinical context, where the presence or absence of symptoms of pathology is used to guide the most appropriate 'treatment'.

Diagnostic testing sessions are conducted one-on-one in a quiet room. The investigator and each participant, on their own, collaborate in a conversation around the use of a particular document. The investigator makes it clear to the participant at the outset that the purpose of the diagnostic session is to find out through the diagnostic testing what, if anything, is wrong with the document. The investigator asks the participant to undertake a number of tasks with the document, recording what they do and say whilst trying to complete the task. Participants are prompted to talk about what they are doing and any problems they encounter.



Three types of quantitative observations are made by investigator with each participant:

- Can they find the information? y/n
- Do they have difficulty finding the information? y/n
- Can they use the information appropriately once they have found it?
y/n

Alongside this quantitative data, investigators report their detailed observations on the types of difficulty participants have in finding information, and report the verbatim comments of participants throughout the diagnostic sessions.

Performance requirements and protocol

In this case, the symptoms of pathology were the unsuccessful or inappropriate uses of Credit Card Statements (CCS) by participants.

With the help of our volunteers, we collected a convenience sample of CCS from around the world. These were depersonalised to remove any information that would identify the specific card holders.

Using the CCS as our starting point, we developed a set of performance requirements for this type of document. Performance requirements consist of two things: a list of the tasks that we believe people should be able to perform with the document, and an acceptable level at which we expect people to perform those tasks. Usually, when part of a design project, the process of compiling and agreeing to a set of performance requirements involves extensive consultation with all stakeholders. However, in this case, the logistics of undertaking such consultation defeated us, and we had to fall back on our own prior experience with similar documents. Figure 3 shows the performance requirements developed for this study.

Figure 4: performance requirements for CCS

| identification tasks | basic usage tasks | interactive tasks |
|--|--|---|
| Identify what the document is (credit card statement) | Find and explain the date range covered by the statement | Find and explain how to make a payment* |
| Find and identify who is providing the statement (company name) | Find and explain the opening balance | Find and explain how to find more information |
| Identify who the credit card statement is for (name, address, account number) | Find and explain the closing balance | |
| Find and explain the statement period (i.e. monthly statement, annual statement) | Identify the total of any cash advances for the statement period and the interest rate that applies* | |
| | Identify the total of any purchases for the statement period and the interest rate that applies | |
| | Find and explain any interest that has been charged to the account | |
| | Identify any transaction dates | |
| | Find and explain any transaction descriptions | |

Find and explain the overall credit limit

Find and explain any available credit

Find and explain any payments that have been made*

Find and explain any payments due (when, how much, any overdue amounts)*

Find and explain any terms and conditions*

Find and explain how many pages are included in the statement

*Note that not all statements include this information

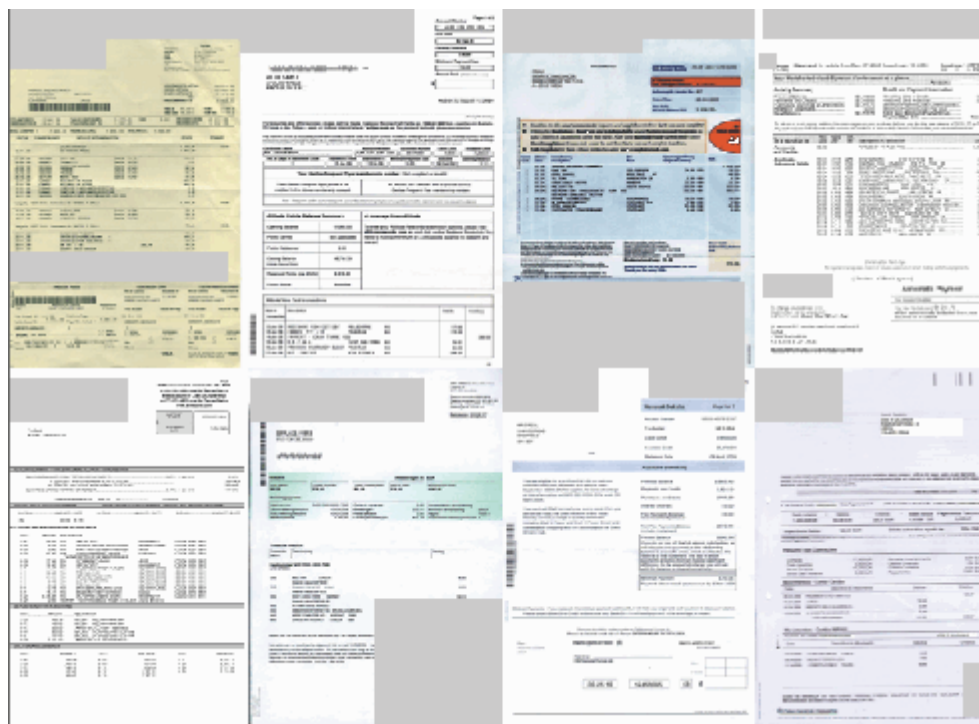
The above performance requirements were then used to develop the test protocol to be used by the investigators in the diagnostic sessions: that is, the list of questions and requests to participants to undertake the tasks specified in the performance requirements.

This protocol was piloted by an investigator in two diagnostic sessions to detect any problems that needed to be resolved before finalising it for use.

Eleven credit card statements from around the world were provided by our volunteer investigators. These were depersonalised prior to the diagnostic

sessions, and are shown in Figure 5 with corporate information blanked out, to avoid drawing attention to particular institutions.

Figure 5: Some of the tested CCS.

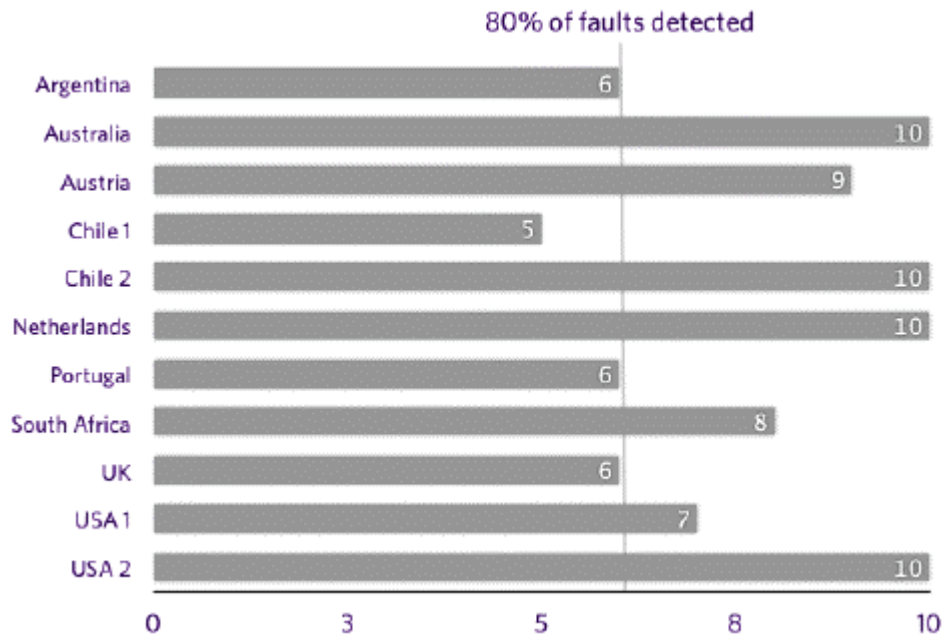


Our investigators followed the diagnostic procedure outlined above, using the same protocol, translated into the local language where needed. They collated the data on standardised spreadsheets and returned them to our project manager in Melbourne, Alex Tyers, who checked them, conferred with each investigator to resolve any queries, and then analysed and aggregated the data. All personal information about the specific participants at each session remained confidential and were not passed on to our project manager. Those data are presented in this paper.

Results

A total of 87 diagnostic sessions were conducted with 11 CCS in 9 countries (Figure 6).

Figure 6: the number of diagnostic sessions conducted in each country

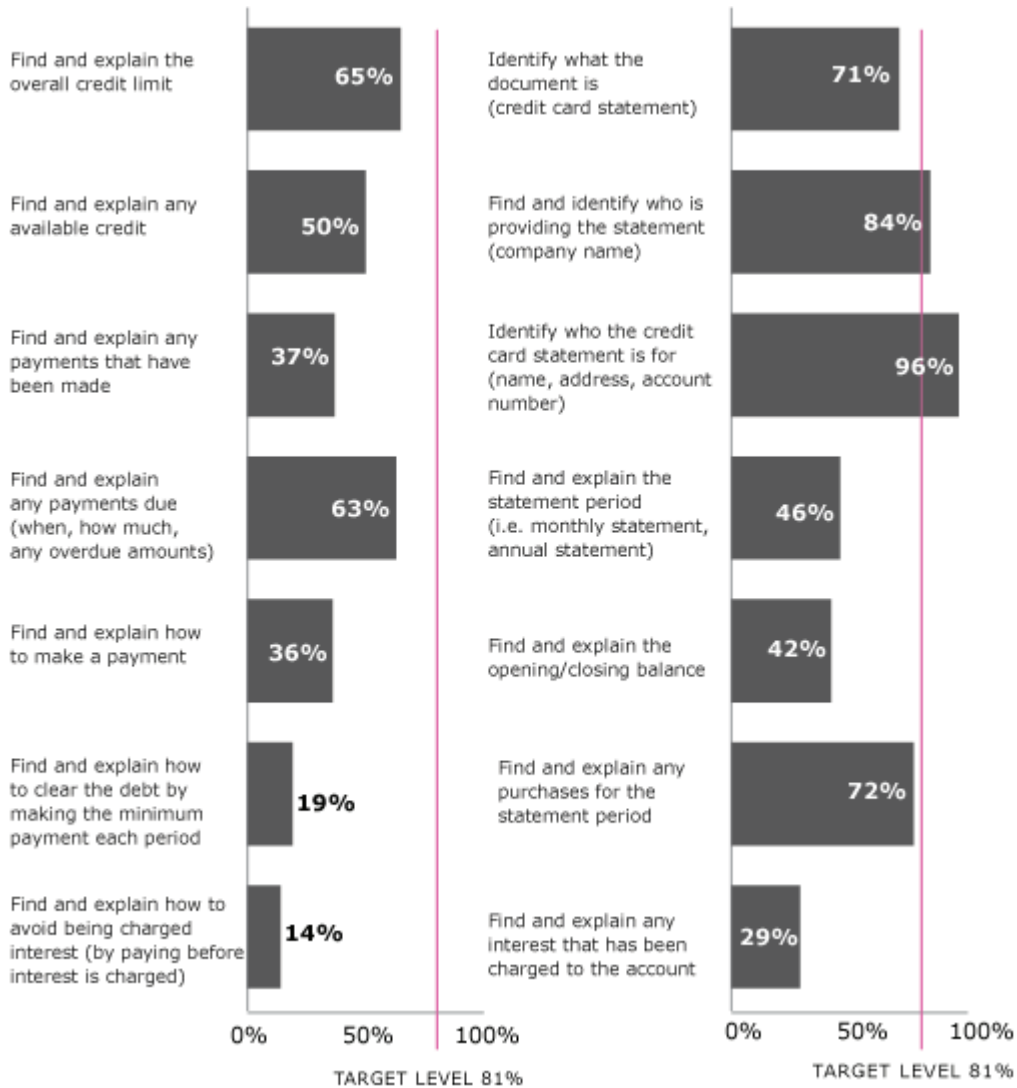


In keeping with earlier benchmark studies we set a target performance level for each of the tasks participants were asked to perform. We used a performance level that has been found to work in many other information design contexts, namely that any literate participant should be able to find at least 90% of what they are looking for on a CCS and then use appropriately 90% of what they find. We used these percentages to arrive at an overall target performance score of 81% by multiplying these two figures together. This composite figure is a 'headline'; it draws attention to the presence of faults in the design. When the components making up these numbers and their related qualitative data are examined together, a full diagnosis of each fault can be undertaken. These overall figures provided us with a picture across all the tasks participants performed and all the CCS that were tested.

The aggregated data for each of the tasks, across all the CSS tested and all diagnostic sessions (Figure 7) show that only 2 tasks out of 14 reached the target performance level:

1. Find and identify who is providing the statement (company name).
2. Find and identify who the credit card statement is for (name, address, account number).

Figure 7: performance level of Tasks across all CCS



Twelve tasks were below the target performance level, some well below.

These under performing tasks ranged from 14% (avoiding interest charges) to 77% (identifying purchases on card for statement period).

Discussion

The three types of quantitative observations made by investigators with each participant (could they find the information, did they have difficulty finding the information, could they use the information appropriately once they found it) leads to simplified scoring which provides useful headline figures indicating the overall performance of a document.

But when taken with the investigators' notes from observing participants' actions and writing down participants' verbatim comments, the result is a detailed story rich with data, much of it providing invaluable qualitative insights into the faults and the reasons for them. These data are extremely valuable not only for information designers helping industry improve their designs, but also for regulators to identify key performance indicators to incorporate into regulations to lift the minimum standards of CCS to an acceptable level. Here we concentrate on the headline figures which are of more interest to the general reader rather than the specialist information designer.

Diagnostic logic

Conventional thinking suggests that the focus of diagnostic sessions is people, that it is people who are being tested. But if we take that view, then we would be required to offer an explanation of the results in terms of people not just their actions but their inner cognitive processes as well. While we can observe people's actions, we have no access to their inner cognitive processes, and consequently we would be involved in a set of inferences based on current cognitive theory not the firmest foundation on which to build an explanation of what is happening in this context. Moreover, we do not have to be cognitive scientists, as we are not in the business of changing people. We are in the business of changing designed information.

Also, if the focus is on the people, there is an implied criticism of them: it is the people who are having difficulty using a document and the implication

is that it is their fault. Most commonly this leads to the easy argument that if people are having difficulty reading, they have a 'literacy problem'. (In many countries the term 'financial literacy' is used as little more than a way of excusing poor document design in the financial sector.) Thus there is no need to redesign a document, because the problem lies in the people's deficiency. Time and time again, our research shows that if there is a 'literacy problem' it is in the organisations producing the documents, not in the people who are the hapless victims of this illiteracy. Blaming the victims does not get to the cause of the problem, nor does it solve it.

The diagnostic logic we follow is to take people's actions with a document as symptoms of the underlying condition of the documents themselves. If a document cannot be used for a particular reasonable purpose, then there is a fault in the document. The pathology is in the document, not the people who try to use it and fail. Moreover, if the document is redesigned, so that it can be used successfully, we take this as evidence that the sick document has been cured of its pathological condition.

Sample size and data quality

We are often asked 'How many people do you test in order to get useful data?'. The short answer, using the above diagnostic logic, is 'None'. We don't test people, we test the information they try to use. This may seem an odd answer, particularly if you come from a background steeped in social science research methods, but the force of this quick answer lies in the way it directs attention away from the study of people to the study of information. We aim to bring about desirable change in everyday information, not, as stated earlier, to bring about change in the people who have to put up with this information.

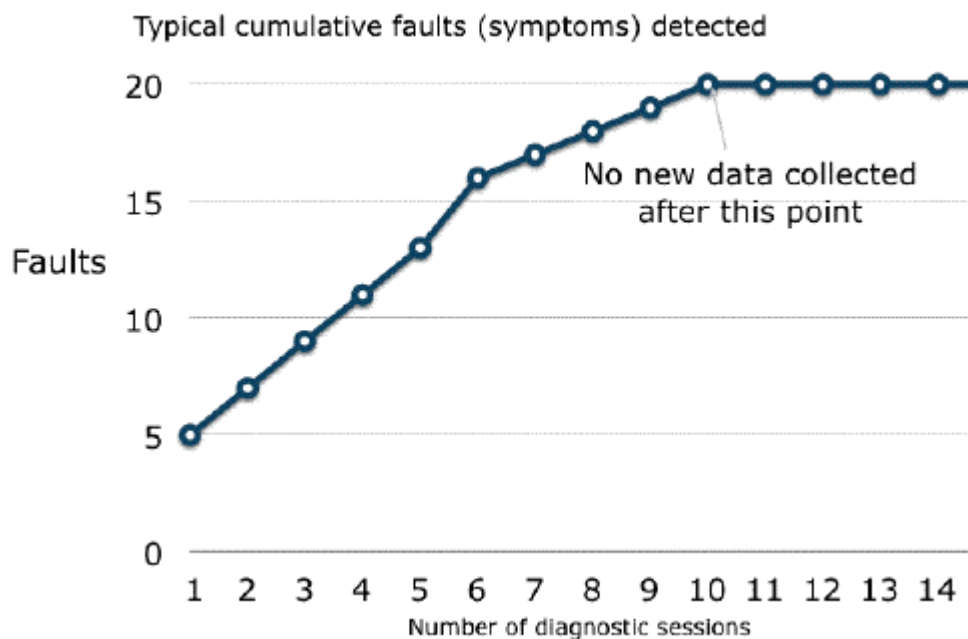
The longer answer is very much tied to what we are investigating, namely the faults in designed information. The question we ask is subtly inflected by this

interest: we ask, 'How many diagnostic sessions do we need to conduct, to identify all the faults in a design?'

Looked at from that point of view, our approach has been to keep on conducting diagnostic sessions until we stop collecting any new data about a design's faults.

The cumulative evidence from research and experience suggests that the first 6 diagnostic one-on-one sessions, each with a different participant, enable the researcher to identify approximately 80% of the faults in a design arising from the tasks participants are asked to perform. After 10 such sessions, approximately 100% of these faults have been detected. So in the eleventh and subsequent sessions, no new data is collected. Figure 8 shows a typical pattern of the cumulative data in such studies.

Figure 8: typical cumulative data on design faults



The percentage of likely detected faults in each CSS is shown in Figure 9.

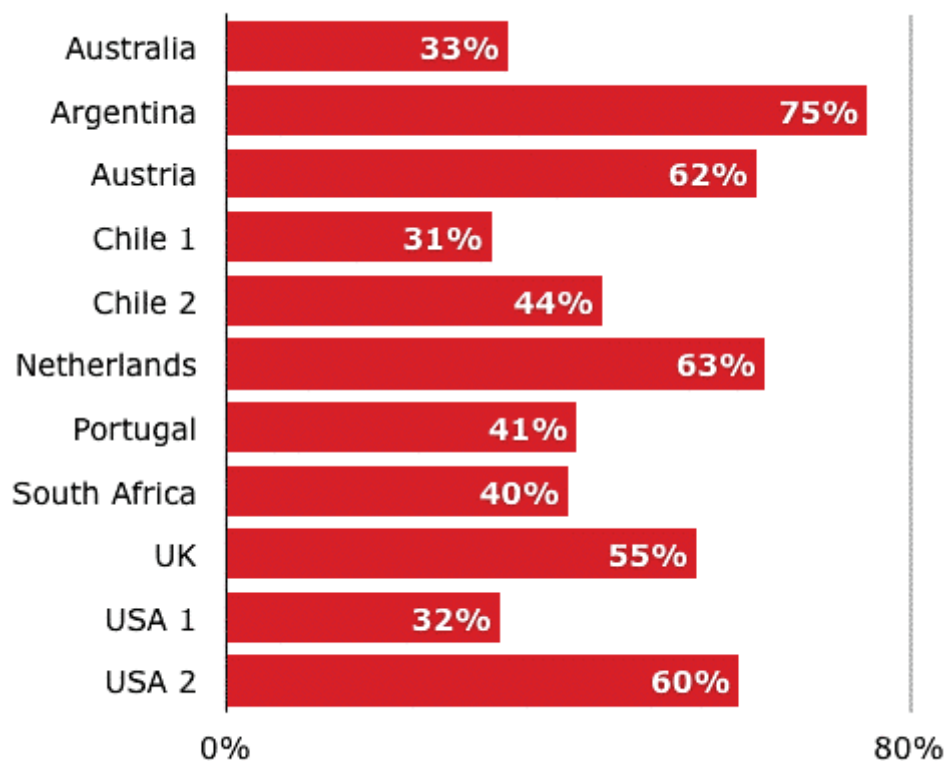
As Figure 6 shows, with the exception of one of the samples from Chile, we have probably captured 80% or more of the faults for the main tasks one would expect the statements to be used for. Of course, there are many more

tasks that participants might have performed for which we have no data, and there are probably far more faults waiting to be discovered in these designs. However, we can be reasonably confident that the data we have collected have identified many major faults in these designs.

Poor performance

We were surprised by the overall poor performance of these documents. We expected that at least some of the 11 CCS studied might achieve an acceptable performance level. But aggregating the data for all tasks performed on each CCS showed that none achieved an acceptable overall performance level of 81%. That is, none of the CCS tested could be successfully used to find 90% of the information, and when found, successfully used on 90% of occasions. Figure 9 shows this aggregated set of results and the number of sessions for each CCS that were at or above an acceptable level.

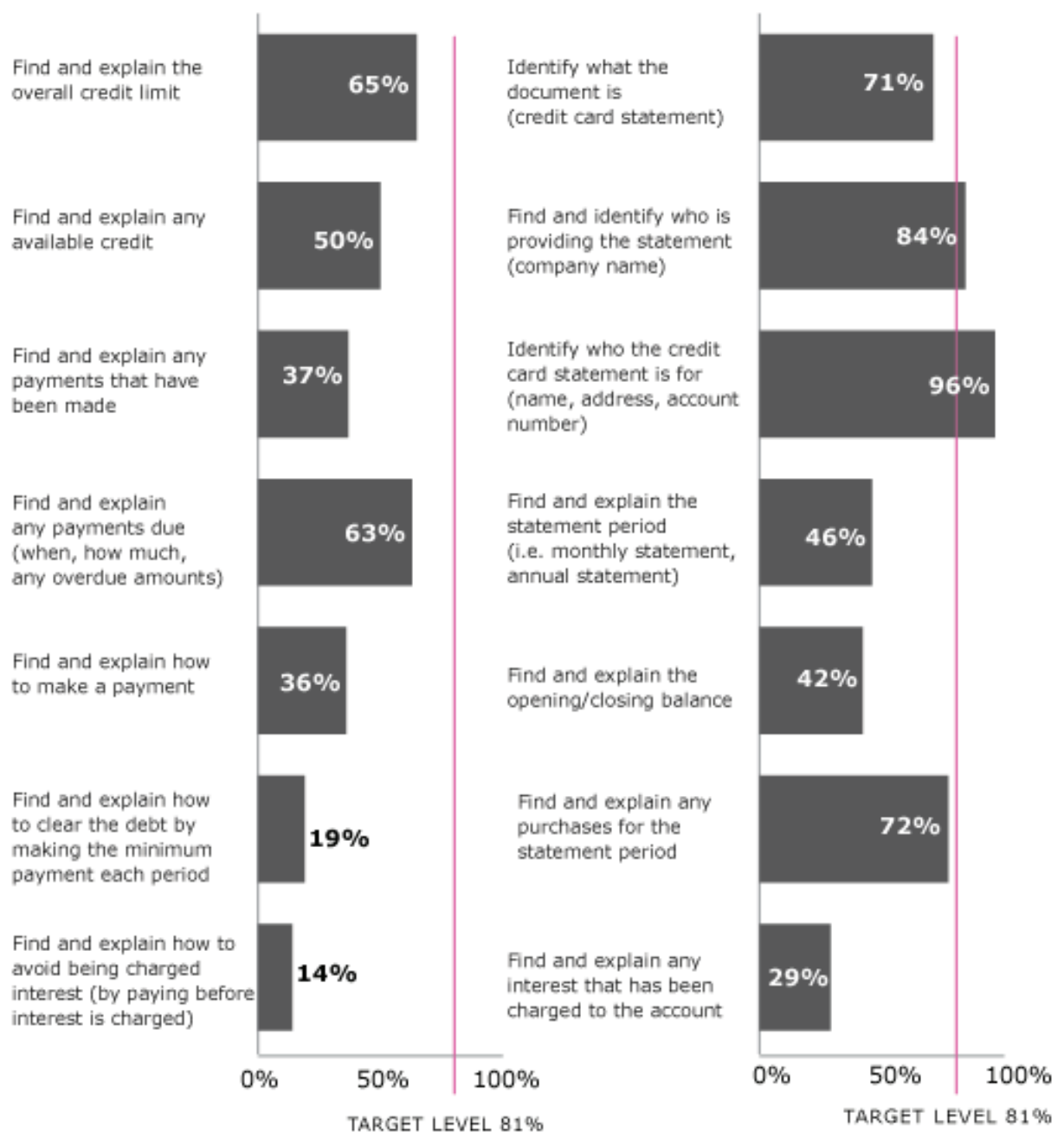
Figure 9: average of overall failure across all CCS



A CCS is firstly an itemised bill, secondly a detailed record of transactions within the bill, and thirdly, it is an account of the business rules applied by the service provider.

The most important information that consumers want to know about a bill is how much they have to pay, when they have to pay by, and how to make a payment.

Figure 7 (repeated) performance level of Tasks across all CCS



As Figure 7 shows only:

- 63% could use the CCS to work out how much to pay
- 27% could use the CCS to work out when the payment was due
- 36% could use the CCS to work out how to pay.

A few consumers will go to the next level of detail and want to know the way in which the various items on the bill are charged. Here too the CCS presents consumers with a challenge. On average, 72% could use the CCS to work out the items that were being charged for (this percentage would probably be higher if they were looking at their own transactions.).

And when it comes to using the CCS to work out the business rules, the CSS provides little help. On average, the CCS could only be used 29% of the time to work out the interest rate that was being charged. In 3 cases this was because the information was simply not there to be found. On average, only 14% could work out how to avoid interest payments, and only 19% could use the CSS to work out the consequences of paying the minimum amount due each month.

Only 65% could identify the credit limit, and, related to this, only 50% could identify how much credit was left, and as a consequence that may not be able to work out what they have spent.

These are the tasks that the diagnostic sessions explored. We suspect that many of the other business rules applied by credit card providers would be equally if not more difficult for consumers to work out in the current designs.

We get a sense of the frustration and irritation for consumers from their comments after using the credit card statements for what should be straightforward tasks:

Figure 10: Participants comments following their attempts to use CCS

It's confusing and not easy to find things. Too many boxes. There's no overall structure hierarchy. I don't know what interest applies. Where are the fees and how much are payments and interest? There's too many boxes, the amount due isn't there and the late fees aren't there. How can I pay?

It even leaves a somewhat seedy impression.

There is a lot of information I don't understand or I don't know what it is there for.

Terrible! It is not well explained, I don't understand the vocabulary they use and the way the sum appears on the top of the list is extremely confusing.

This is not for people who are not used to forms. It's a disaster.

You should not buy such a card. It is a useless statement. It is only there to confuse you.

(The layout) is very confusing, too much information, often repeated, leading to doubts.

Bad. There are 40 things on here and I can't find anything easily! (It does not tell you) how I can pay; when a payment due date; and what the statement period is.

It is in such disarray – to me it looks like a high school project.

Sh*t! You don't know what a lot of it means.

Confusing! It doesn't clarify any of my spend, intimidating, doesn't give me information that I'd need.

Not easy to read, lot of information, even though it's got bold, I automatically think it's not going to be easy to understand. I don't like tables – It is too hard to follow the lines. I can't read across them.

It doesn't give you what you want to see. You have to search for what you want to know. It is very unclear as to whether you have made your payment and the interest charge you would incur.

It is more difficult than I thought and visually overwhelming. I'm glad my spouse deals with these bills.

I would love to see information on how to reduce interest rate and state very clearly where customer service can be accessed ridiculous!

It feels like a deliberate withholding of information and obfuscation, I would like to see a clearer more upfront, visually.

It's really confusing, made worse by the fact that I use a statement like this all the time.

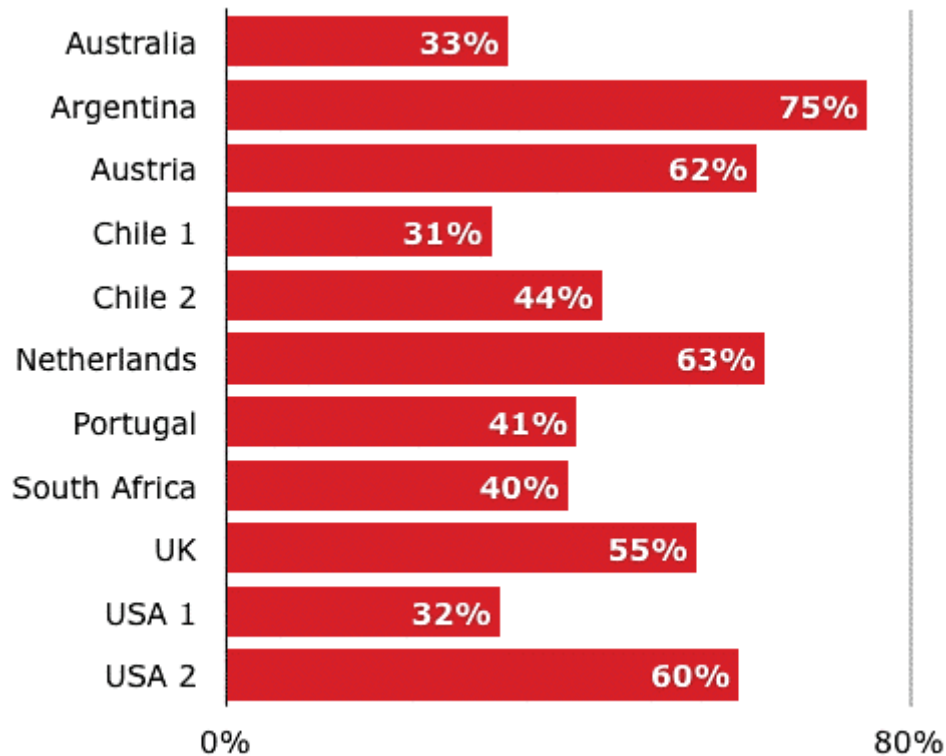
The final comment sums it up:

It's a f*****g nightmare!

Conclusion

The picture to emerge from these findings is one of systemic failure. This is most tellingly illustrated by aggregating the data across all tasks for each of the statements tested. Not one of them gets to the acceptable target performance level of 81% (See Figure 9).

Figure 9 (repeated): average of overall failure across all CCS



It is tempting to see this systemic failure as a symptom of conspiratorial action by credit card providers. However, to do so would require us to ascribe a degree of wilful dissembling and deliberate engineering or design of the documents to make them unusable. This would require the Credit Card Providers to have at least some skills in sophisticated information design, and there is absolutely no evidence of this in the designs we tested. Indeed, these documents look like many others to emerge out of contemporary information factories, not through a process of deliberate design, but as the end product of amateur typography and a lack of systematic and rigorous information design processes.

More likely, then, these documents and the pathological symptoms they display are the result of uncaring neglect. Insofar that this neglect provides cover for some unacceptable business practices, regulators need to take firm measures to protect consumers. But, based on our experience, we would advise regulators to specify the tasks that customers should be able to

perform with the documents and the acceptable level at which they should be able to do so, leaving the execution of particular designs to professional information designers. The current practices in some regulatory bodies to specify both the content and appearance of a document with the same low level of skill as is currently applied by industry in creating these documents will assist no one, least of all the consumers, and an industry with predatory intentions will use their compliance with the letter of the law as a new form of cover for some new predatory practices.

By specifying the tasks (and leaving the information design to enable those tasks with industry), the room is left open for innovation and for market forces to provide incentives for good design. As we have seen in other industries, businesses which are first to market with new and innovative designs can capture a significantly increased market share, and the less innovative then copy the winning designs. In the end the customer benefits.

We were disappointed that this Communication Benchmarks study found such uniformly poor designs, and we want to encourage industry to do better in the future.

However, we would not recommend any of the opportunistic suggestions by graphic designers. These are highly speculative sketches not based on any benchmarking data, nor have they been tested. As the evidence from many previous studies suggests, such speculation is rarely an acceptable solution, and may not even be a good starting point.

We would like to repeat this particular study in 2010, when companies have had an opportunity to see these results and learn from them, and also after they have had time to respond to some of the newer regulatory requirements for this type of document. We would like industry to offer us their best examples for the next Communication Benchmarks study and we hope we can at that time publish a happier set of numbers.

References

[1 Banking website benchmark studies](#)

[2 Medicine labelling benchmark data](#)

3 Clarke T, Tyers, A, & Sless D (2003)

Testing financial service guides Financial Products, Financial Service Guide
REPORT TO: INVESTMENT AND FINANCIAL SERVICES ASSOCIATION
Canberra: Communication Research Institute of Australia

[4 Government and business forms case histories benchmark data](#)

[5 Utility bills benchmark data](#)

[6 Consumer Medicines Information \(CMI\)](#)

7 There is an article describing the reasons for choosing diagnostic testing [here](#). A full account of how to conduct diagnostic testing is contained in [Writing about medicines for people](#), and in our [Professional Guidelines](#).

8 Beime B, Lawrence M, Rieke C (2007)

The practical application of readability user tests in national and international marketing authorisation procedures
Regulatory Rapporteur May Issue 2007 8–13.

9 Faulkner, L. (2003)

Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments, & Computers 2003, 35 (3) 379–383.

Remember this is a summary of the 'headline' data collected in this study.

If you would like access to the full data and our analysis of what needs to be done to improve specific credit card statements, please contact [David Sless, our Director](#).

If you are interested in participating in our Communication Benchmarks Projects as a volunteer, or in any other capacity, please contact [Alex Tyers, our Communications Benchmarks Manager](#).