

A UK Usability Professionals Association guide -

Key questions to ask your usability testing supplier

How do you know if your usability supplier really knows what it's doing?

More and more service providers (e.g. agencies, contractors, etc.) are offering what they call 'usability' or 'usability testing' – but the quality of these services varies greatly. Hiring a supplier that gives misleading, or unhelpful advice is a waste of time and money – and can mean that your new system (e.g. website, handheld application, software solution etc.) is no better than the old one.

To get the best value for money, you have to ensure that your service provider takes the best approach to usability testing.

This checklist will help you choose a usability-testing agency

The UK Usability Professionals' Association is committed to improving and maintaining standards in the usability community. This document gives you some of the key questions you should ask your usability supplier and explains what to listen for in their answers. You can use it as a checklist during invitations to tender, when interviewing usability providers or when reviewing proposal documents.

This checklist covers choosing a supplier to help with **usability testing** – but usability experts are important in defining requirements and creating designs. In the future, more advice will be available from the UK Usability Professionals' Association on these topics. In the meantime, some of these questions may still be useful when you are talking to suppliers about those aspects.

The questions

These are the key questions the UK Usability Professionals' Association thinks you ask any supplier you use:

1. What usability techniques are appropriate for this project?
2. What standards will you follow and what measurements will you take?
3. What users will you test?
4. Will I get helpful and accurate answers?
5. How 'usable' will your deliverables be?

The rest of this document summarises the things you should listen for in your providers' answers – and explains best practice in a little more detail.

1.0 Use the most appropriate techniques

There is no one ideal method of gathering usability inputs. The optimal method will depend on the stage of the design process you are currently at, and the nature of the results you need. Most good agencies will be able to recommend a range of techniques, and discuss the relative benefits of each, in relation to your circumstances.

1.1 Interviews, surveys and focus groups

We don't recommend any of these techniques for testing usability. These approaches are good for gathering users' opinions, understanding their problems or getting their perceptions of past behaviour. Whilst all of these insights are of interest to usability professionals, they are not meaningful measures of *usability* – just users' opinions and beliefs.

1.2 Expert reviews

The quality of expert reviews depends on the level of experience of the expert and the review technique used. They can be useful for quick usability assessments – but they don't identify as many problems as testing with real users. It is advisable to do an expert review before a real-user usability test – to pick up obvious problems or to help plan the test.

There are a number of expert review techniques that are widely used and understood (such as Heuristic Analysis, GOMS, NGOMSL, GOMS-KLM, Cognitive Walkthrough, Norman's Seven Stages of Action). We recommend using these, rather than an undocumented technique – that way if you're not happy you'll be able to find someone else to review the results.

1.3 Testing with real users

There are a number of different approaches to real-user usability testing. These can range from highly diagnostic and analytical to measurement of system performance to meet targets.

For diagnostic testing, we recommend tests are conducted one-on-one, with an interviewer observing a user who uses the system to complete meaningful goals. It is common that users are encouraged to narrate their thoughts as they progress - usually known as "Thinking Aloud Protocol". In most cases, the observer will only work with one user at a time – not spreading their attention thinly between a large number of users.

Some tests are aimed at *measuring* user performance on a given system, in order to guarantee usability targets. It is possible that these types will allow multiple users to participate simultaneously. Your supplier should be able to discuss and advise on the benefits of different approaches.

Checklist: Will your supplier use the most appropriate methods?

- The supplier can discuss the relative benefits of a range of methods
- The supplier can give examples of previous successes of different methods
- The supplier can distinguish between different types of real-user testing and advise accordingly
- The supplier offers a range of expert review options

2.0 Acknowledge measurements and standards

A good supplier should know how to *measure* usability. By measuring usability each time you test, you'll be able to see whether or not your system is improving.

The most basic measure of usability is 'what proportion of users completed the tasks they were set'? However, there are a number of other important factors. Professional suppliers will be able to work to ISO (International Standards Organisation) standards.

A commonly accepted definition of usability (ISO 9241-11) covers 'effectiveness' (degree of task completion), 'efficiency' (how much effort it takes to complete a task) and 'satisfaction' (positive attitudes toward the product and freedom from discomfort). Suppliers should be able to suggest ways of measuring these things through observing users.

Other common measures are 'learnability' (the effort and resources required to learn the system) and 'memorability' (how well users recall how to use the system after a period of time). Learnability and memorability matter most to frequently used systems such as intranets worksystems. Often, they're regarded as aspects of 'efficiency' and 'effectiveness'.

A good supplier will also point out that measuring something only makes sense if you have a basis for comparison – perhaps an alternative design or a competitor's website. In all cases, measurements should be meaningful – arbitrary scores such as "the site is usable for 60% of users" won't help, unless "usable" is unpacked, and 60% is compared to some benchmarks.

Depending on your needs, you may want to ensure that diagnostic analysis doesn't come second place to usability measurements. If you are looking to redesign rather than simply evaluate performance then a finding of "80% users failed to complete this task" won't help without also telling you *why* and *what you should do*.

Another ISO standard that the supplier should work to is ISO 13407 – the way to integrate usability inputs into an ongoing user-centred development process.

Checklist: Does the supplier acknowledge measurements and standards?

- The supplier can discuss the appropriateness of measurements
- The supplier can work to recognised ISO standards
- The supplier uses meaningful and actionable measurements, rather than just "numbers for numbers' sake"
- The supplier can discuss the limitations of measurements, in terms of statistical significance

3.0 Test with the right users

Any testing is better than no testing – you will always learn something you didn't already know, and it opens people's eyes to how users really behave. However, to ensure you are using your resources most effectively, you must ensure the right users are participating.

Firstly, the service provider must recruit the right users. This, as ever, depends on the nature of your project. However, you should be able to ensure that a number of groups are represented:

A range of demographics that match the profile of your audience

Key usage segments such as existing users, first-time users, users of equivalent systems, churned users

It is always important that the users employed are not "test professionals" or work in design industries – they will not reflect the natural usage that a system affords in the real world.

Secondly, the service provider must employ the right number of users. You will be paying more for every user, so you need to be sure that they are all adding value to the outcomes.

There is no standard number of users for a successful test. It depends on your needs, the number of different segments that are of interest, and the complexity of the system in question:

'I need to identify usability problems'

Assuming all your users fall into a similar category, the right facilitator will require perhaps five to ten users to identify most of your problems. If you've got two distinct groups of users, then you'll need to double the number tested, and so on.

'I want to get a guide to the performance of my website'

If you want to get reasonably accurate usability metrics to base design decisions on, you'll need at least 15 users per group. This is usually enough to show you how much of an improvement you can expect from your new designs.

'I need to predict how my system will behave'

If you're trying to *accurately* predict out how a new site will affect sales or productivity, then you need to test yet more. If you want figures that are accurate to within about 5%, you'll need to test 40-50 users.

So, although you don't need to test with many users to *identify* problems, if you want to start *measuring* performance, you'll need to test far more.

Checklist: Are you testing with the right users?

- The supplier can provide a screening documents for user recruitment, and justify their inclusions
- The supplier understands the importance of sample size, in relation to your needs
- The supplier can describe the statistical implications of larger sample sizes

4.0 Give helpful and accurate answers

Good usability is not the same as people *saying* they like your designs. What users *say* when observed often bears little resemblance to how they behave when actually using the same system.

Ask suppliers to provide previous reports and case studies of prior outcomes. If reports concentrate on what users *say* about systems, or how they feel about certain functions, it is likely that it will give you no accurate basis for design decisions. If the report concentrates on aesthetics or branding then it will not help you create more usable systems. Equally, if reports contain arbitrary figures such as “usability is 8/10” or “users were 75% happy”, then you need to challenge the supplier on their underlying thinking. The findings should focus things like *where* and *why* users were not able to complete their goals, and *where* and *why* users became frustrated.

When it comes to getting a test subject to use your system, we recommend setting general scenarios, rather than specific tasks, where possible. A scenario would be ‘you’re thinking of buying some flowers for Mothers’ Day – use this website to help you’; a task would be ‘send a bunch of ten lilies to your mother’. Scenarios allow the user more freedom of action (including the freedom to give up) and so are better approximations of real behaviour.

Users often *will* give useful verbal insights into the system. However, the best way to elicit this is for the facilitator to ask open ended questions such as “what do you want to do now?”, “what can you see?”, and “what do you think will happen when you do that?” and not questions like “Do you like this screen?”.

It is also important that the person facilitating the test does not offer clues and cues for the users. For instance, if your website has a link called ‘products’, facilitators shouldn’t ask users to ‘look for a product’ – they’ll pick up on the cue in your question, and your results will be of limited use.

In some cases you will want pure “summative” evaluations and reports of your system. However, it is often the case that your usability efforts will be with redesign in mind. If this is the case, ensure the provider can move beyond stating where users experienced problems, and begin to describe *why* this occurred and *what the solution is*. A good project will result in actionable findings and recommendations – though for certain solutions, responsible agencies should want to prototype and re-test before they are confident the problem is solved.

Checklist: Will you get helpful and accurate answers?

- The supplier focuses on demonstrated problems and not users’ feelings
- The facilitator does not ask closed questions or lead users
- The supplier uses meaningful scenarios to test systems
- The supplier can work to your technical and commercial constraints, where appropriate
- The supplier can suggest actionable solutions, not just state problems

5.0 Offer useful and fit-for-purpose deliverables

Even if you have an accurate and insightful set of findings, it is no guarantee that your stakeholders will understand or even acknowledge them. The delivery of findings must be excellent, and suitable to your particular circumstances.

There is no such thing as the perfect set of deliverables. They will range in terms of inclusions, format of presentation, and the degree of support that the supplier offers in their communication and uptake.

You know best the nature of your organisation and which deliverables are likely to fit within its processes. It may be that development teams have the operational capacity to work through a detailed log of findings and recommendations. Alternatively, they may respond best if senior management are offered a high-level personal presentation of results first.

Nevertheless, some of the typical features you should expect from usability deliverables are: detailed observation log with severity indications; illustrations of findings - either static annotated screenshots or videos of the problem occurring "in situ"; Quotes and comments from users; Analysts' interpretive diagrams and comments; Quantitative analysis of usability measurements across the user sample; graphs and charts reflecting this same information; summaries of priority problems; and in some cases visual representations of solutions and recommendations.

It is likely that beautifully illustrated reports are not optimal for rapid, repeated testing and design iterations. The accuracy and helpfulness of the results should take precedence over the form of their delivery.

The deliverables will probably be most useful if they are presented in increasing level of detail: an executive summary, the main (analysed) report, and the raw data or observation logs.

Very compelling presentation formats can be used in order to ensure your organisation fully buys into the findings: picture-in-picture video evidence of user interactions tend to achieve this.

Agencies will also differ in the degree they support the communication of their deliverables. Some will be happy to simply deliver standalone written documentation. It is likely that you will also look for a personal presentation of findings. More sophisticated suppliers will be keen to workshop the interpretation of their inputs, perhaps over an extended number of sessions. Anything beyond this might well be considered user-centred design services, as opposed to usability testing services.

Checklist: Will you get useful and fit-for-purpose deliverables?

- The deliverables can be tailored to meet your needs
- The deliverables contain analysis, illustrations, raw data, and recommendations
- The deliverables are easy to digest: findings are prioritised and grouped meaningfully
- The supplier can support your team in the understanding and uptake of findings