

Enabling International Usability Using Multicultural and Geographically Disperse Teams

Elisa del Galdo¹, Sushmita Munshi², and Christine Truc Modica³

¹ Chief of Technical Staff, Human Factors Europe, UK
elisa.delgaldo@humanfactors.com

² Project Manager, Human Factors International India
sushmita@humanfactors.com

³ Design Strategy Lab
christine@designstrategylab.com

Abstract. This paper explores the challenges of running large scale international usability tests and attempts to provide useful information on how to effectively and efficiently design and administer these types of projects. The information in this paper is based on the experience of the three authors, who have all been involved in a number of large-scale international usability tests as managers of the projects, designers of the tests and as practitioners responsible for the collection, reduction, and analysis of user data. Practical advice is provided on how to best approach the many challenges of running these types of projects.

Keywords: Internationalisation, **Internationalization**, Usability Testing, Remote Usability Testing, Cross Cultural Design, Multi-national Teams, International Project Management, International Project Logistics.

1 Introduction: Addressing the Challenge

In today's global marketplace, businesses are looking for ways to efficiently and effectively develop products, systems, and services that are not only desirable to foreign markets, but also acceptable in terms of their usability and applicability for each local market. These same organizations are also looking to lower their costs of development and increase their profit margins to an international market. As corporations strive to offer truly international and easily localisable products and services, the affect on the usability community is clearly visible. The amount of international and multinational usability tests being performed and the increase in number of practitioners worldwide are clear evidence. So how do we, as usability specialist, effectively design and efficiently implement international user testing, such that we offer a quality service at a reasonable cost?

For smaller single location and large international usability companies alike, international user testing can present a number of challenges. The first of these is the logistics of running an international test. Previously straightforward tasks for one country can become an onerous and complicated exercise when having to hire

facilities and equipment, organize recruitment, hire local moderators and translators, and schedule participants in an unfamiliar country. The organization of many of these activities will fall on the overall project manager, and may prove to be especially difficult when done from a long distance and in a foreign language.

Another issue is the design of the tests themselves. Practitioners will need to determine if the usability testing methods they intend to incorporate into their study will be an effective way of data gathering in all locations. This means that it may not be a matter of simply designing one test and translating the test document for other regions. The entire test methodology may have to be altered to suit the characteristics and norms of the local participants. Then there is the problem of consistency in testing techniques and methodology. How do you ensure that your tests are then consistent and collecting like-for-like data? There is also the cost of having to design and manage the implementation and potentially different data types from multiple methods.

Also to be considered is the provision of moderators and consistency between them for multi-location tests. Do you use the same moderator for all of the countries and provide simultaneous translation for each language session? Using the same moderator for all of the tests will certainly increase the consistency among the tests, data reduction and analysis. It will also increase the time span of the testing and can significantly increase the cost of testing due to the need for translation services and travel.

This paper explores the requirements of running large scale international usability tests and attempts to provide useful information on how to more effectively and efficiently design and administer these types of projects. The information in this paper is based on the experience of the three authors, who have all been involved in a number of large scale international usability tests as project managers, test designers and as practitioners responsible for the collection, reduction, and analysis of user data.. Practical advice is provided on how to best approach the many challenges of running these types of projects.

2 Project Management and Logistics

The first thing to work out when beginning an international project is the project management. Working in multiple locations can prove to be a highly interesting learning experience and as well as something of a logistical nightmare. From the outset of the budgeting of the project to planning and administration of testing, the project manager is required to have detailed knowledge of each testing location and the elements of the local culture that may have an affect on usability testing. Being aware of and planning for the factors detailed below can save a lot of time and effort for the project manager and also help to keep to a budget.

2.1 Creating Project Budgets

When planning and budgeting international tests, it is important to start with time and budget constraints and then select relevant usability methods and a participant sample that fits well within these constraints. The collaboration with local partners can help you plan a relevant budget based on the countries and user groups you want to target.

It is important to remember that the cost of moderators, facilities, equipment rental, and translators (to name a few) may vary greatly depending on the countries in which you are testing. This, along with other unexpected costs, such as requirements of travel stipends in some countries on top of a daily rate and expenses, or the requirement of translation of a moderators notes before data reduction, must all be documented in the budget calculations. If the project covers several countries in different continents, it is good practice to add 10- 25% on the top of the calculated budget to cover any unexpected costs..

2.2 Recruiting Participants

The number of participants and location of the usability tests should reflect the diversity of the target user groups. European and Asian markets have particularly wide-ranging cultures and languages. Ensure your locations reflect the diversity of your user groups.

When doing recruitment across several countries or cultures it is important to remember that a descriptive term, such as ‘trend setter’ may not always apply to the same demographics in each location. Also, the description of a social strata that holds true in one culture, doesn’t necessary hold true in another. For instance ‘family income’ in the UK means the income of a couple, in India it means the extended families cumulative income, and in Japan it is taboo to ask a person’s income at all. Thus it is recommended to research societal standards in different cultures or to hire help of local experts before organizing recruitment of participants.

The same is true with the location of the recruiting. In some countries, you find differences in behavior and customs between people living in the capital city and people living in other cities. For example, in the biggest US cities, people with high incomes tend to live in the suburbs; this is generally not true in France, the UK or Italy. Knowledge of local customs can help avoid recruiting mistakes and bad sampling.

When scheduling tests, be aware of bank and summer holidays (the months of May or August in France can be challenging for recruiting). It may be very difficult to recruit the right people during holiday periods and you may face a higher rate of cancellation of those recruited.

2.3 Renting Testing Facilities, Equipment and Services

The cost and quality of testing facilities and services can vary from country to country. Most commercial hubs have testing facilities with appropriate equipment to conduct and record the usability tests. If your target user groups are not located in a big city or a commercial hub, it is wise to rent a conference room in a comfortable hotel or business centre located centrally. It may be necessary to bring your own equipment. Avoid conducting usability tests in regular hotel rooms as this is normally not acceptable in many countries.

If you do decide to use your own equipment, ensure that it will work well in each geographic location and be prepared for all eventualities. When testing in cold climates (when it is the norm for even indoors to be very cold) be aware that some electronic equipment may not work well or at all. The same would be true for very

hot climates where there is no air conditioning available. And always be sure that you have plenty of plug adaptors for each location and spare batteries. Always visit the testing facility a day before the usability tests so you have time to take corrective action in case of problems.

2.4 Translation of Test Documents

For all local language tests it is important to contextually translate all key test instruments. (e.g., Recruitment Screener, Test protocols and moderators guide). Literal translation can completely alter the purpose or meaning of a question. Back translating (translating the document back into its original language) protocols is recommended to ensure your translations are of high quality, but can increase the cost of translation. Pilot testing and having your local moderators check the documentation will highlight any problems or inconsistencies in the translation or problems due to cultural specific differences.

2.5 Legal Requirements for Consent

Legal requirements for testing can be very different from one country to another. This is particularly important when usability tests involve minors, skilled labor, or patients. You should check and review legal regulations and rules with a local recruiting company, then work with the local moderator to develop relevant consent forms. In cultures with high power distance, participants are often taken aback or are uncomfortable signing long legal documents. For example, while testing with illiterate migrant workers in India a number of participants refused to sign consent forms or were very nervous. By Indian law, verbal consent is a legal alternative, so in this case it was not necessary to get a signed consent form.

2.6 Recruiting Moderators and Translators

It is always best to use moderators and translators that have a background or experience in usability testing. When using translators and local moderators it is important to ensure they understand the purpose and of the usability test. The competence and understanding of moderators and translators is vital as language issues might otherwise prevent problems being identified. One technique to minimize confusion is to pilot the test using the moderator or translator as a sample user, so that they better understand the process and aims of the test. It is recommended to make sure that the translator and moderator in question have adequate domain knowledge to comfortably translate complex technical or domain specific jargon.

2.7 Establish Local Partners

Along with the recruitment of moderators and translators locally, selecting and consulting local partners at the beginning of the project can help to avoid planning and project management mistakes. It can be time consuming to work with local partners, but in fact they can save time and money in the end. Local partners can help you understand the local culture and people and to solve planning and development problems you might encounter.

3 International Teams for Moderation of Tests

The whole world is moving away from the ‘we design they consume model’. Increasingly clients want input for their designs from local usability professional and they want the input early in the development cycle. Using international teams to conduct multiple location usability tests is necessary to get this information quickly. In many parts of the world the usability profession is nascent. Therefore finding readily available, trained and experienced usability professionals may not always be easy. A combination of a small core team and dispersed local teams can prove beneficial to the smooth running of a project. The core team provides over all direction, methodology and quality control, while local teams execute tests at the ground level and can provide insights into cultural differences that a non-local usability professional cannot.

The key benefits of using local usability teams are:

- Local teams provide valuable local and cultural insights which go un-noticed to outsiders.
- The use of international teams allows for quicker turnaround of projects as testing in multiple locations can happen simultaneously.
- Time zone differences ensure that someone is always working.
- Overall cost of using local partners is lower than using a single global team as travel costs are negated.

Clear communication is the backbone to the success of any project. Lack of a common language and differences in cultural norms are one of the biggest hurdles to overcome in international projects. Even if the whole team speaks the same language it does not ensure smooth sailing. A large number of professionals comprehend written English better than they are able to speak English. Providing all communication in written form or following up verbal conversations with the details repeated in emails can significantly assist in communication. It is recommended to have key requirements of the project documented, as the written word is more easily comprehended and recalled than the spoken word.

At the onset of the relationship it is necessary to communicate the purpose of the test and roles and responsibilities of everyone. This knowledge empowers local teams to make intelligent suggestions and take decisions in your absence. Time zone differences can also be an irritant; as it is difficult to find convenient times for meetings or phone calls. The project team may have to alternate convenient times for conference calls in order to avoid any one location feeling less important than others.

3.1 Ensuring Consistency and Quality Control

How do you ensure consistency in the test methodology and in the data reduction? The whole engagement is always easier if there is a common language between local moderators and the overall project manager, but this may not always be possible. There are various ways to ensure consistency of data and its reduction across locations, given below are some points to keep in mind.

- Sharing templates, exemplars and timelines early on in the engagement ensure expectations are communicated.

- Often a reverse engineering process helps. First create a format for your final report, this helps you understand the different kinds of data that needs to be captured. The protocol and other testing instruments are optimized and aligned to the final report format. These documents when shared with the larger team ensure the correct data is captured.
- Ensure that all documentation is translated.
- Ensure moderators, translators and observers view recordings of sessions to familiarize themselves.
- The core team should review data logs of dry runs and provide feedback to local teams. These ensure inconsistencies are ironed out before you go live.

4 Test Design and Methodology

4.1 Test Documentation and Translation

If at all possible the test documentation should be translated by the translator who will run the test or by the moderator. If the test documentation is translated by a person who will not be involved in the test, then it should be reviewed and edited by the actual moderator. This will ensure that the usability test sessions will run smoothly. The objective of each task should be clearly presented in the documentation in order to give the moderator the ability to ask the same question in a different way if the participant doesn't understand the question or if the question doesn't solicit the information intended. Also note that formal, I demonstrative languages increase test time by about 25% more time than in the UK or US. It is important to take this into account when planning and conducting a test.

4.2 Design and Prototype Translation

Users might not share a common perception of geography as the developers, for example Asian countries and Australia are listed together under APAC. Users usually do not perceive them together. Icons, metaphors, colors can be full of meaning in a country and meaningless in another based on the way people work and live. It is important to take into account these localization issues when translating designs or prototypes. Conducting a pilot test with the moderator or translator should help to identify some of these issues. It is important to note that the purpose of the usability test should be to identify these types of issues that are more abstract in nature. But, localisation details that are more commonly recognized, such as formats of dates, times, and addresses should be addressed in the translation or localisation of any test material.

4.3 Methodology vs. Culture

The way people react to a usability test can be very different based on their age, level of education and culture. When working with children we often customize standard usability test methods. We have to do the same with users from other cultures and be creative and open to find the best way to fit their cultural behaviors.

For example, it is usually very difficult for Indian users to ‘talk aloud’ during a usability test. Immersing Indian participants via a scenario seems to help take attention away from the ‘seriousness’ of the task. Thus allowing them to articulate the problems they encounter with the system being tested without the fear of feeling inadequate or offending the designer.

In Asia generally, Users will have difficulty to criticizing a design, even they don’t like it. They consider that it is not polite to do so. In Europe, people will have the tendency to over criticize and in Africa people will use only the extreme such as good or bad. It is important to take this cultural difference into account when the test includes satisfaction rates.

4.4 Reduction and Translation of Data

Based on the scale of the project, this phase can be a very lengthy. To make this work easier, it is very important to plan and structure this phase in advance with local moderators. Templates should be developed to describe the type and level of feedback the moderator needs to record. Results linked to cultural specifics should be highlighted.

As an example for all moderators, the pilot test results could be summarized and translated and then be used as a reference document for doing the same with the actual test data.

4.5 Analysis of Data

Be aware that for some data it will not be possible to compare across cultures. For example, comparing user ratings to judge levels of satisfaction can often be misleading. As noted above, users in some cultures are more critical than in others, so absolute ratings (e.g., using Likert scales) may show great variation between cultures due to different cultural norms.

5 Alternative Methods for Small Budgets

Budgetary constraints are a part of usability testing reality. When funds are tight the validation or iteration activities are one of the first activities to be reduced or cut out all together. Such knee-jerk reaction can prove to be expensive in the long run. The importance of iteration and validation from end users cannot be stressed enough, but often clients may not see the necessity of this process as they already have one set of results. One way to circumvent this situation is to use alternative testing methods that are less expensive as a way to validate the results of smaller one-on-one usability testing or increase the geographic range of the tests.

5.1 Moderated Remote Usability Tests

Remote usability testing (RUT) is a reliable and cost effective method to gather user data and iterate early and finished interfaces with a geographically varied user group. As screen and voice sharing technologies are more reliable, remote usability tests are now practical and easy to execute.

The key difference between moderated remote usability tests and Lab based tests are:

- The moderator, respondent and observer are at different locations
- A software tool or application allows the moderator to view the participants screen.
- The moderator and participant communicate usually via a conference call. VOIP often proves to be cumbersome and increases the time lag. Using the phone is also practical if the internet fails, the moderator will still have a phone to guide the participant through reconnection.

5.1.1 Benefits of Running a Moderated RUT are:

- Recruiting benefits
 - Reach geographically dispersed populations
 - Reach hard to recruit populations – participants do not have to leave their home or office
- Maximize returns on a small budget
 - A larger group of participants can be reached with a limited budget.
 - Travel and lab costs are reduced.
- Quick turn around time for results. Less time is spent on planning and scheduling thus overall a larger number of tests can be conducted in less time.
- Performance based tests, test of self evidence; brand perception tests are extremely effective when administered remotely.
- Ethnographic benefits. Participants are tested or interviewed in their actual environment

5.1.2 Key Disadvantages of Running a Moderated RUT are:

- Recruiting and Logistics: RUT's ensure one can reach a wide variety of user group. But as it is remote it also means the control on quality of recruitment is lower.
- Dependence on technology
 - Screen and voice sharing devices can often prove extremely unreliable.
 - Non-technically savvy user groups are awed or uncomfortable by the RUT set up and thus do not perform optimally.
- Dependence on moderator's skills: RUT's are heavily dependant on the moderator's ability to connect with the participant. Often if the moderator fails to build a connection with the participant, the data proves to be light and without any key insights
- Qualitative, contextual, formative tests are tough to execute via an RUT: The geographic barrier makes it tough to capture facial expressions, body language, and environmental cues. The personal contact of in person usability tests is lacking in a remote usability test and participants often do not openly critique or discuss issues.

5.1.3 Points to Remember While Planning a RUT:

- Identify the key purpose of running an RUT and the main questions which need to be answered. The protocol design needs to be optimized to answer the questions.

- Protocols designed for RUT are tighter and more direct than an in-person protocol. Un-structured, bloated protocols are culprits for failed tests. Validate the protocol in advance. Blindfold a colleague and read the protocol aloud. Quiz the colleague to determine if they remember the flow of the questions. This will throw up obvious bottle necks which were not so obvious. The aim is to ensure that the protocol addresses the purpose of the test and at the same time is easy to comprehend.
- Seamless technology that is fool proof is necessary to successfully run remote usability tests. A dedicated and reliable internet connection, a good telephone line to transfer voice, and a reliable and easy to use screen share application is necessary. When taking technology decisions ensure the participants will have the infrastructure to support your technology requirements. Ask yourself: “Do factory workers in Krakow have dedicated internet connection to support a heavy screen sharing application?” If not, then RUT may not be feasible for your intended user group.
- Control and cross check recruitment: To ensure you get the correct sample, develop a detailed and if possible localized screener for your recruitment agency. Ask the agency to send you a spread sheet which contains answers to screener questions for each participant. Cross check the appropriateness of participants at the onset of each test session. Do not feel uncomfortable rejecting participants, but be aware that these participants will still need to receive remuneration.
- Experienced moderators are imperative for the success for a RUT. In case the tests are in a language that you don’t speak. Find a moderator who speaks the language and train the person. Using a two-way translator is a good strategy but it almost doubles the test time. In case of multi language testing write your protocol in non metaphoric language. As direct language is easier to translate and is not open to interpretation.
- Consider time zones and culture: Most RUT’s are conducted across time zones; plan well in advance so that neither you nor your participants are forced to conduct tests in the middle of the night. People from different cultures react differently when being interviewed remotely. Some become more open others retreat and become more formal. Thus take cultural and social characteristics into consideration while analyzing the data.

5.2 Un-moderated Remote Usability Tests

Un-moderated usability tests are the cheapest way to validate a design and reach out to a wide variety of user population on a very limited budget and resources.

5.2.1 Benefits & Drawbacks

- Cost advantage: Un-moderated usability tests are the cheapest way to validate and reach out to a wide variety of user population.
- Un-moderated remote UT’s work best in scenarios where the designer already knows the attributes of the target user group well and the user group is fairly homogeneous. One can send out the test to as many people one wants, as the cost of mailing is minimal and remuneration is considerably lower.

- Flexibility of time and location: Participants can take the test as and when they require.
- Easy to plan and execute as there are fewer middle men: No moderators, translators or observers.
- It is often impossible to collect qualitative behavioral insights. Surveys, binary choice answers, card sorts, brand perception are most effective.
- The protocol design needs to be tight and focused to ensure participants do not mis-interpret the questions or run through them in a disinterested manner.

The two most popular ways of running a un-moderated RUT are

1. Test instruments with instructions are mailed to the participant. The participant completes the test in his/ or her own time and mails it back.
2. The test can be hosted on a secure website. Participants are invited to take the test. The participant's inputs and actions are captured on the site.

Both these methods have their pros and cons. The latter in recent years have proved to be more reliable than the former. Primarily in the case of tests which provide contextual question and help because participants feel that there is someone at the other end cross checking the data. Both these methods usually have between 10-40% no shows or dead data. Thus it is imperative to send out invites to a large participant group.

It is also important to ensure that the participants are comfortable and with the method used. Ask questions: Does my user group have dedicated internet connection to take this test? Are they comfortable with the media we are using?

5.3 Mixing of Traditional UT and Remote Testing for Validity

In the last few years using a combination of traditional and remote usability tests has become more common. Usually in person tests are conducted at the onset of a project and RUT's are used to quickly iterate a proposed design structure in the latter half of a project. This strategy is recommended as a good alternative, especially as the freedom to change the design reduces dramatically in the latter stages of the development process. But, the reverse, doing an RUT first and following up with an in person test, is often not very beneficial. In the case where both methods need to be used in the same round of testing, test using traditional methods first and then use the remote method to check your insights or hypothesis with a larger group of participants.

6 Conclusion

This paper has provided useful information on how to successfully plan and administer international and multi-location usability tests. The most important guidance is to ensure that each usability test is well planned and designed to accommodate the requirements of testing all of the user groups. As with all usability tests, the thorough planning and meticulous preparation are the key to the collection of good quality data and the problem free running of tests. Understanding all of the additional factors involved in the running of international tests will help the practitioner to create a sound plan and run successful usability tests.