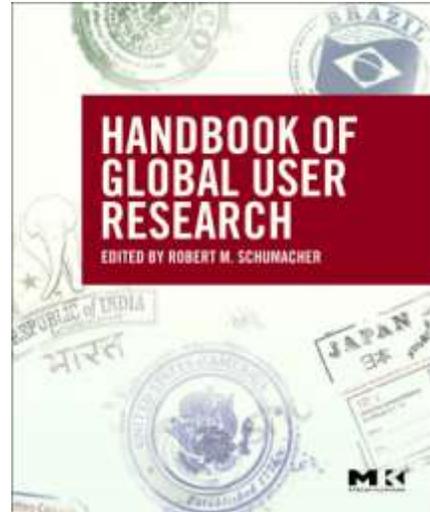# 3. Preparation

Agnieszka (Aga) Bojko, *User Centric, Inc.*

Jessica Buttimer, *User Centric, Inc.*

Sokol Zace, *ErgolineUSA*

3.1. Introduction

Proper preparation is essential to the success of any research study. Bosenick, Lew and Darriba in Chapter 2 provided background information on planning (e.g., locating resources and staffing a team), here we turn planning into plans and detailed preparation. This is especially true for studies that that involve multiple countries, cultures, languages, and research teams. Preparation for global studies involves three processes:

- comprehensive study planning,

- local team training, and

- creation of backup plans.

Comprehensive study planning for a global project includes typical study planning activities, but at a greater level of detail than a single location study might require. The main goal is to make sure that the study answers the right questions building on a foundation of high quality data. The lead research team (i.e., the team that has a relationship with the study stakeholder) has to understand the research objectives, review the stimuli to be studied, determine the appropriate methodology for data collection, and create the testing and recruiting materials.

The second process of global research preparation, local team training, involves communication with teams in other geographies. The objective is to ensure proper and consistent recruiting, data collection, analysis, and reporting. This communication can take the form of a detailed Test Plan with pictures of the lab setup, additional explanations of the procedure in the Moderator's Guide, sample video from the study conducted in the lead team's country, and a debrief following an on-site pilot session.

Even the best of plans can go wrong, so backup plans must be set in place. Being prepared for the worst case scenario is always important, but in a global study, it is crucial. For example, it can happen that the devices to be tested do not make it through customs or that the Internet connection necessary to access the online stimuli is down. Postponing the study is often not an option, so alternative solutions should be at hand.

In this chapter, we focus on global aspects of research preparation and present them in approximate order of their execution during a project, as shown in Figure 3-1. We discuss the following topics:

- o Understanding the research objectives and target user groups
- o Reviewing the stimuli and making sure they work in all test locations and are properly localised
- o Creating the key document of any project – the Test Plan
- o Recruiting participants, including screening, scheduling, and compensation
- o Developing research materials such as the Moderator's Guide for user testing
- o Localising the research materials
- o Sharing research materials with the local teams

- o   Briefing of the local teams

- o   Training local teams through internal or external pilot testing

We conclude the chapter with a number of key take-aways based on the insights, tips, and examples discussed in the chapter.

<Insert Figure 3-1 About Here>

Figure 3-1. An example of the order of preparation activities in a three-country study, where data is first collected in the lead team country (location A), followed by parallel data collection in locations A and B.

3.2. Understanding research objectives and target user groups

The first step in any research study is to identify the key research objectives, thus enabling appropriate approach and methodology to be employed. In a global study, it is particularly important to understand how the test locations were selected, what the stakeholders hope to learn from each country, and what their expectations and concerns are for each country. There are several ways to choose countries for a study, for example:

- o   The top countries with the highest market share for the artifact studied

- o   The top markets from each continent of interest

- o   The top markets from each language group of interest

- o   Developing markets

- o   Problematic markets

- o   A combination of the above

Based on their communication with the study stakeholder, the lead research team should determine what cultural aspects are being studied. Is the study focused on the differences among the locations, similarities, or both? As much as we may sometimes be curious to pursue country- and culture-specific findings, common themes across the locations may be more valuable to the stakeholders. Focusing on the common themes is especially important if only one version of a product or interface can be created, and high degree of customisation, beyond basic localisation, is not feasible. For example, the primary goal of global usability studies for existing products is usually to make sure that the product is fundamentally usable, regardless of the users' culture or location. Recognising issues specific to each location or culture is also valuable but secondary.

If the research objectives involve particular product brands (e.g., competitor products in a comparative study), the market share of these products and their brand perception in the selected countries should be thoroughly understood, as they may impact both recruiting and testing. For example, certain mobile phones that are common in Europe may be used by very specific and difficult-to-find user groups in North America. If that is the case, the lead team should provide feedback to the study stakeholder, so either recruiting timelines can be extended or project scope can be adjusted to reflect local reality.

In addition to establishing clear and achievable research objectives, the study stakeholder and the lead team should define the target user groups that should be represented in the study. As the first step, they should determine if the user profiles are valid across countries and cultures. Basing user profiles on demographics alone may be misleading, as the meaningful similarities across locations may be goals, needs, or other less tangible characteristics. Therefore, it is often beneficial to create international personas when preparing for a global study. More information

about personas and detailed steps on how to create them are provided by Nielsen in de Boer's chapter (Chapter 6) - Personas in a Global World.

3.3. Reviewing the stimuli

3.3.1. Taking care of localisation issues

Once the objectives are clear, the next step is to become familiar with the product or interface of interest and understand its capabilities and constraints. In a global study, the research team should pay particular attention to localisation issues. Unless an existing product is being tested or the focus of the study is on how well a product has been localised, poorly localised stimuli can produce skewed or unclear results.

For example, if Western imagery is displayed in a prototype of a Web site tested in Asia, participants may not be as positive towards the site as they would be if it featured Eastern imagery. What complicates matters even more is that participants may not be willing or even able to verbalize why they cannot relate to the site. Another common example is an improperly translated link or button label, which can prevent participants from accessing the task-relevant section of the interface, thus causing an unnecessary failure. In this case, unless participants are told directly where to go, it may be impossible to gather data on the usability of the further steps in the task.

Ideally, the local teams should review the translation of any prototype stimuli, and necessary changes should be made prior to testing. If there are known issues that cannot be fixed before the study, the lead team needs to decide how these issues should be handled during the study to

avoid data loss. For example, the moderators can be provided with a list of acceptable hints they can give to participants who stumble over an improperly translated task-relevant term.

The lead team should also verify the availability of the stimuli in all languages that are needed, even if the stakeholders believe that all language versions exist. It may be too late if, for example, the device about to be tested in Germany turns out to be available in Dutch rather than "deutsch" (the German term for "German"). Another important consideration involves language differences between countries that speak "the same language". For example, Portuguese used in Brazil is different from Portuguese used in Portugal, and certain technical terms commonly used in mobile phones may not make sense to the Brazilians if they are asked to interact with a phone in European Portuguese.

Problems can arise when the research stimuli are already preloaded onto laptops and cannot be used on any other computer, as keyboard keys and layouts are not the same in every country. There are several different keyboards for Latin scripts (e.g., QWERTY, QWERZ, AZERTY, QZERTY), non-Latin alphabetic scripts (e.g., Arabic, Hebrew, Russian), and East Asian languages (e.g., Chinese, Korean, Japanese). Even if the countries of interest to the research use, for example, the QWERZ layout, there are still several variations of it, based on the location of symbols and special characters. Even keyboards used to type in the same language can differ. For example, the keyboard layout that Austrians use to type in German differs from the layout used for Swiss German.

Because asking participants to type on an unfamiliar keyboard may affect the results of the study, the lead team should investigate how different the layout is from a local keyboard and determine if the differences are relevant to what participants will be asked to do. Pictures of all available

Windows keyboard layouts can be found in the Microsoft Go Global Developer Center (msdn.microsoft.com/goglobal) under Learn > Tools and Utilities. If a test location may be affected by an unfamiliar keyboard layout, it is best to connect the laptop to a monitor and a local keyboard. Attaching a different keyboard to the system may require a change to the setting for the keyboard type in the computer software.

3.3.2. Making Sure the Stimuli Work

In addition to taking care of localisation issues, the teams should make sure that the stimuli are compatible with the facilities used for the study all test locations. A sufficient number of appropriate adapter plugs and voltage converters (see sidebar for more details) should be secured prior to the study. If the study uses mobile phones that need service, then it is also important to check network coverage and availability of particular mobile services in each location. For example, if participants are asked to surf the Web on the phone, the phone will need a data package. While prepaid pay-as-you-go SIM cards can be easily obtained in Europe, they are not as prevalent in the US and may not include features such as data plans.

If the study stimulus is a consumer application that connects to a telephone line, it is important to know whether it uses an analog or digital line. For example, most residential homes in the US have analog telephone lines, while many businesses, including focus group facilities, have digital lines. A dial tone simulator should be obtained if an analog device is tested on a digital line. Appropriate phone jack adapters may also be necessary for studies where devices need to be connected to phone lines.

Ideally, the local research teams should have access to the stimuli well before data collection, so issues can be identified early on, and there is enough time to secure any necessary additional

equipment. If the stimuli need to be mailed to the local teams, the lead team should do it as early as possible, in case they get held up at customs. However, even if the stimuli are not available ahead of time, some basic research (e.g., checking the strength of a 3G mobile signal at the test location) can still be useful.

<sidebar>

Adapter plugs vs. voltage converters

An **adapter plug** is a device that enables the connection of a power cord to an electrical wall outlet that has plug holes of a different shape, number, or arrangement than the plug. An adapter plug does NOT change the voltage of the power source. 64% of countries use one outlet type, 30% use two, and 6% (e.g., Iraq, Kenya, and Singapore) use three. Examples of adapter plugs are shown in Figure 3-2.

<Insert Figure 3-2 About Here>

Figure 3-2. A variety of adapter plugs.

A **voltage converter** is a device that changes the voltage of an electrical power source. Most single-phase alternating-current electrical outlets in the world have either a voltage range of 210 to 240 volts or 100 to 120 volts. Converters usually halve or double the voltage to make it suitable for electrical devices made to work in the other voltage range. Each country tends to have only one voltage range (78% of countries have the higher voltage and 15% have the lower) but some countries (e.g., Algeria, Korea, and Peru) have both. Examples of voltage converters are shown in Figure 3-3.

<Insert Figure 3-3 About Here>

Figure 3-3. Two different voltage converters.

A comprehensive list of plug and socket types, voltages, and frequencies can be found in Wikipedia under "AC power plugs and sockets."

</sidebar>

Sometimes, even after the stimuli have been confirmed to work, they may fail to function during data collection. For example, the Internet connection in India can go down during the monsoon season and that can affect studies that require online access. A good backup plan will ensure that the study can be conducted according to the timeline. If the research stimulus is a Web site or a Web application, the local teams should have a copy that can run locally on a computer. Even using a hard copy of the interface screens can provide valid data about the usability of the interface. Also, in case the power goes down, computers and other devices should be connected to battery power units, so they can stay on for a while and the sessions will not get interrupted.

3.4. Creating the Test Plan

A Test Plan is a document that helps organize all information about a study. It is a great communication tool between the lead research team, local test teams, and the project stakeholders. A detailed plan will ensure that all local teams have a clear and consistent understanding of the study and can start planning and preparing before the recruitment and test materials are ready. The Test Plan should be continually updated as changes occur and more details are discovered or clarified.

A good Test Plan contains:

o Objectives of the study broken up into specific questions that the study has been designed to answer.

o Description of the stimuli, including version or model numbers where appropriate (the same products can have different model numbers in different countries).

o Description of the target users with an indication of strict requirements, allowable variances, and criteria that may need localisation.

o Description of the methodology, including the procedure (e.g., think-aloud, within- or between-groups design) and measures used (e.g., time on task, ratings, rankings).

o Requirements for equipment needed for testing (e.g., size of TV or computer screen, computer operating system, types of cameras).

o Pictures/sketches of the lab setup, denoting how the stimuli should be arranged, where the moderator and participant should sit, and where the camera(s) should be set up. These pictures are especially beneficial if the setup is complex and includes multiple elements (e.g., three or four different camera shots and several artifacts).

o Screenshots of expected video output and acceptable video formats.

o The project schedule, including test dates, times, and locations in every country as well as deliverable deadlines for stakeholders, local teams, and the lead research team.

o Contact information of all team members in each country, their time zones, and their availability, which is especially important if team members are working on multiple projects or if testing is taking place near holidays.

Ideally, the Test Plan should be written in a language with which all local teams are familiar. Alternatively, it can be created in the lead team's native language and then translated into a common tongue, such as English or Spanish. If not all teams have a language in common, the Test Plan should be translated into their respective languages. Regardless of the language, the Test Plan should be comprehensive and concise. Bulleted lists and diagrams help convey important details in a clear manner. Convoluted sentences, long paragraphs, and unnecessary words and phrases make information more difficult to process and should be avoided. Figure 3-4 shows excerpts from two different Test Plans – one for a Web site study and the other for a comparative mobile phone study.

<Insert Figure 3-4 About Here>

Figure 3-4. Excerpts from two different Test Plans.

3.5. Recruiting

3.5.1. Screening

The next step in study preparation is recruiting. Based on user profiles, the lead research team creates a screening questionnaire (*aka* "screener"), according to which the study participants will be recruited. The screener questions should be clear and unambiguous, as too much latitude can lead to poor participant selection.

Once the screener has been created, it needs to be translated into the local languages of the test locations. The translated questions and recruiting criteria must be further adapted to the locations in which they will be used because countries and cultures differ in many ways. Levels of education, race descriptions, and household income ranges for social classes are some of the basic differences between locations. Also, certain job titles can represent different job functions. For example, a "Manager" in the US may have the same responsibilities as a German "Operator."

Quantitative screening algorithms (i.e., procedures that feed candidates' answers into formulas which calculate numerical scores) that work well in one location can be impossible to use in other locations because the same score can have different meanings in different countries or cultures. For example, the score that indicates a "low-tech user" in one country can indicate a "high-tech user" in a less technologically developed country. Localisation of an algorithm requires high sample size market segmentation research, which should be conducted well before usability research begins.

Even the number of participants to recruit may need to be adjusted based on the location of testing because no-show rates (i.e., percentages of participants who do not show up for their interviews without cancelling) can vary significantly. In certain locations (e.g., some northern European countries), rarely will participants not show without a notice. In other locations (e.g., some Arabic countries), on the other hand, it can be a common occurrence. To determine how many extra participants should be recruited to meet the quota (i.e., the required number of participants from each population segment), the local teams should provide the lead team with their typical no-show rates for the targeted participant profiles.

Local teams should also be in constant communication with their recruiters and forward key updates to the lead team, whose task is to centrally monitor the recruiting progress. Recruiters should be given a spreadsheet template that accounts for all the information about the participants that is of interest. If the recruiters use the template to enter the recruited and terminated participants' responses to the Screener questions, the lead team will be able to detect any discrepancies between the intended meaning of the questions and their actual comprehension. If such discrepancies exist, any necessary adjustments to the Screener should be made immediately.

3.5.2. Scheduling

When determining the test schedule, there are a number of considerations that the lead team should keep in mind. First of all, the researchers need to find out if there are any holidays or other events (e.g., Carnival in Brazil, Diwali in India, Olympics) that take place at or around the time of the planned data collection period in the countries of interest. Those times should generally be avoided when planning test dates. If there are holidays or special events prior to testing, longer recruit times should be built into the timeline. A list of each country's holidays can be found at a number of locations on the Internet including, earthcalendar.net or in Wikipedia under "List of Holidays by Country."

Once the lead team has decided when the data collection should take place in each location, the next step is to set a session schedule. In some locations (e.g., in northern Europe or Japan), participants tend to come in early or on time. In others (e.g., some Latin American and southern European countries), time tends to be less rigid and it is culturally acceptable for research participants to be late, sometimes even up to 30 or 45 minutes. Fortunately, these participants usually do not mind staying later either. The test schedule in countries where time is perceived as

flexible should be more relaxed with longer breaks in-between sessions than in the countries where punctuality is more highly valued.

If the schedule needs to be changed at any point during preparation, the impact of rescheduling participants can vary across countries. For example, participants in the US tend to take the recruiter less seriously if their test sessions get rescheduled more than once. As a result, the show rate dramatically decreases. However, in other regions (e.g., southern Europe), participants' behavior is not affected if their appointments get rescheduled.

3.5.3. Consent Forms, Non-Disclosure Forms, and Compensation

Participants' reactions to consent forms and non-disclosure agreements vary across cultures. Some participants may be hesitant to sign a document containing a lot of legal language. Therefore, forms should be written in simple language and include only the necessary information. Local research teams can either localize the original form provided by the lead team or use their own after adding the details of the study. Local teams should also be aware of the legal requirements in their country and make sure that the documents fulfill them.

As in any study involving participants, compensation depends on the studied artifacts and the target user groups. For example, to test a recently launched high-end mobile phone in a developed country, the remuneration should be higher than typical. Alternatively, other country-appropriate compensation forms (e.g., dinner for two in an upscale restaurant) can be used because people with a high socioeconomic status are not easily convinced to participate in studies.

Because the "market value" of the same user groups varies from country to country, test locations will differ in the amount of participant incentive. There are also differences in the

method of payment. For example, in some countries (e.g., China, UK, US) cash is used for most user groups, while in others (e.g., Finland, Japan, Spain), gift cards or other forms of compensation such as, for example, magazine subscriptions are more prevalent. The non-cash incentives are considered to be "more elegant" than cash and are often used to minimize taxation issues.

The local team in each country should be able to determine the proper form and amount of participant incentive. For participants who are prohibited from taking any compensation directly, the local team can make a charity donation on their behalf.

The presentation of consent forms and compensation may also differ across locations. In some locations, participants are handed the consent form in the waiting room prior to the session and are paid after the session either by the research team or by the recruiter. In other locations, participants sign the consent form and are compensated by the study moderator at the beginning of the session. When there is no impact on the results of the study, it is usually best to give the local teams freedom to do what they normally do, rather than impose the procedures employed by the lead team in their home country.

<sidebar>

Observed similarities and differences in recruiting activities between the US and Italy

Recruiting activities	USA	Italy
Sourcing candidates	Ads (online and in papers) attract large numbers of "fresh" candidates. Cold calls work well too but take a lot more time.	Few people answer ads because many Italians have not had direct experience as market research participants. Most assume the ad is a camouflaged sale or fraud. Cold calls work slightly better than ads

		but candidates tend to be suspicious because of their previous experience with unethical sales people. The best way to source fresh candidates is to call your social network including previous participants and ask for referrals. It is often necessary to relax the screening criteria, because of the limited number of candidates available.
Pre-screening via brief email survey	Email communication works well. Answers tend to be clear.	Email communication does not work as well as in the US because answers can be unclear.
Interviewing	There are sophisticated liars who have made a profession out of participating in research studies. However, the majority of interviewees are honest.	There are no sophisticated liars because few people have participated in user research before. Liars can be easily identified and exposed.
Making the appointment	Participants tend to check their calendars to make sure that they will be available before making the appointment.	Some participants make the appointment without a careful evaluation of their previously scheduled engagements.
Calling with a reminder	Participants appreciate the reminders but most already have the appointment in their calendars and would not have forgotten.	Reminders help identify those who had not checked their calendars when making the appointment. Calling two days prior to the session is better than the day before because participants who forgot about the appointment and made other commitments will have more time to readjust their schedules. Also, the recruiter will have more time to find a replacement if the participant cannot make it.
Predicting no-shows	Most no-shows are predictable as a result of the reminder call. Candidates who do not intend to show up either do not answer the call or sound unmotivated.	No-shows are less predictable based on the reminder call. Some poorly organised candidates may have last-minute situations that will prevent them from attending the study.
Compensating participants	Paying cash at the end of the research session is the best option. Checks are less preferred.	

| Following up | Participants appreciate the follow-up call and tend to give spontaneous referrals. They are also likely to spread the word afterwards. |

</sidebar>

3.6. Developing the Moderator's Guide

A good Moderator's Guide (*aka* "guide") for a global study should be easy to translate into other languages. Therefore, abbreviations, unnecessarily rare terms and phrases, and idioms should be avoided as much as possible to reduce ambiguity. The guide should also be explicit and detailed to ensure correct and consistent data collection procedures across locations. Besides questions and task instructions that the moderator has to say to the participants, the guide should include other information that will help local moderators understand the purpose of the tasks and questions, task priority (e.g., which tasks can be skipped if there is no time for all), required depth of probing, and allowed latitude of probing. The information for the moderator can be inserted in appropriate places in the guide in a way that is easily distinguishable from the main content of the script, which is shown in Figure 3-5. This information should also include stimulus-related instructions, such as on which Web page each task should begin or how to reset the test devices prior to each session.

To finalize the Moderator's Guide, the lead team should conduct one or more pilot tests. Pilot testing will help refine the wording, order, and priority of tasks and questions, determine proper time management strategy, and make sure that the guide is in perfect alignment with the tested artifact(s).

<Insert Figure 3-5 About Here>

Figure 3-5. Sample page from a Moderator's Guide.

3.7. Localising the Moderator's Guide

3.7.1. The Importance of Formal Localisation

Once the Moderator's Guide has been finalised in the original language, it needs to be localised to all test locations. Some local moderators who are proficient in the original language (e.g., English) may be able to moderate in the local language using the original version of the guide. This is *not* recommended, however, because real-time translation can have a detrimental impact on the precision and consistency of data collection. Instead, a fully localised version of the Moderator's Guide should be developed for each country.

Even if the study is conducted in another country that uses the same language as the guide's original language, the document may still need to be adapted to regional differences and dialects. For example, a Moderator's Guide used in Spain needs to be localised before it is used in Argentina because Spanish-speaking countries differ in vocabulary, and common words in one country can confuse or even offend participants in another. Because localisation involves not only translation but also cultural adaptation, scenarios, tasks, and questions may also need to be modified to make sense for participants in different countries.

The Moderator's Guide can be localised either by local teams or by professional translators. Each of these methods has its advantages and disadvantages, which are discussed in the following sections. Regardless of the selected method, the translator should be a native or near fluent speaker of the language into which the guide is being translated. The lead team should make sure that the translator knows what needs to be localised. If the moderators understand the original language of the guide, localising only the parts that will be spoken to the participant will

be less costly and time-consuming than localising the entire guide, including moderator instructions.

3.7.2. Localisation by Local Teams

Engaging the local moderators in the task of localising the Moderator's Guide increases their involvement in the study and improves preparation. The moderators should interact with the stimulus while localising the guide to ensure a close correspondence between the guide and the stimulus. For example, if the guide mentions terminology that appears in the tested interface, the moderators should look it up in the localised version of the interface. This is preferable to independently generating the equivalent terminology in the local language because the terminology may not necessarily match what is used in the interface. Sometimes, to avoid leading questions, the terminology in the guide should *not* match that of the interface. The moderators can then make sure that the terms in the guide are synonymous with but not identical to those used in the interface.

Localisation of test materials by local teams only works if the moderators are familiar with the original language of the guide. However, even if the moderators are perfectly bilingual, the localisation of the guide may take more time and be lower quality than if it was done by trained professional translators.

3.7.3. Localisation by Professional Translators

Localisation by professional translators tends to be more exact, faster, and less costly (because it takes less time) than localisation by local research teams. However, translators may not be familiar with the language used in the user research field and may use phrases that can sound awkward when talking to a participant. For example, the expression "there are no right or wrong

answers" can usually be translated in more than one way but not all of the possible translations may be appropriate to say during a user research study.

There are established language practices for these types of phrases, and ideally, the researchers should find an experienced translator who is familiar with user research or market research. The translator should also be informed of the setting in which the guide will be used. For example, interviews with teens in video game stores require a more casual language than interviews with adults conducted in hospitals or financial institutions. Regardless of how skilled the professional translators are, local teams should review the translated material prior to data collection. If necessary, the local teams should adjust the language to fit the particular user research setting and ensure that stimulus-specific vocabulary matches (or does not match) the tested artifact. Issues caused by an incorrect localisation of the Moderator's Guide may be difficult to detect and overcome once data collection has begun.

3.7.4. Reverse Translation

Reverse translation requires the guide to be translated into the target language by one translator and translated back into the original language by another translator. The resulting document is then compared to the original and any discrepancies are investigated and corrected. Reverse translation greatly reduces the uncertainty that the Moderator's Guide matches the intentions.

Because a reverse translation increases time and expenses involved in study preparation, a cost-benefit analysis should be performed before the decision to conduct a reverse translation is made. For studies that require a very high degree of precision in data collection, such as quantitative validation studies or research in the medical field, a reverse translation procedure may be appropriate or even necessary. On the other hand, a formative usability test of an entertainment

Web site may benefit very little from a reverse translation of the test materials compared to the cost involved.

3.8. Sharing Materials with Local Teams

There are a number of materials that the lead team of a global study will share with the teams in the all test locations. The Test Plan is usually the first document that should be available to the local teams. The Screening Questionnaire and participant spreadsheet template should follow, so recruiting can start while the lead team may still be working on the Moderator's Guide and other materials. Once the guide becomes available, the local teams should be provided with the stimuli, so they can become familiar with the procedure in the context of what will be tested.

If the study requires interpreters, they should be able to review the Moderator's Guide ahead of time as well. Assuming that the study sessions will be simultaneously translated into the original language of the Moderator's Guide, it is a good practice to give both guide versions to the interpreters – the one in the local language and the one in the original language.

When data collection begins (ideally, the lead team's country would be first), and once the lead team is comfortable with how the procedure is being executed, they should select a session and make it available on video to all other locations. This is especially helpful if the local teams understand the language of the video, but even if they do not, a sample video can still be useful. The opportunity to watch a session will significantly increase consistency of data collection across locations.

At some point during the preparation phase, the lead team should also share the data sheets and report templates with the local teams, so the output from all test locations is exactly as needed

and consistent across locations. Data sheets and report templates are described in more detail in Chapter 5 on Analysis and Reporting.

While sharing electronic test materials via email works well for smaller projects, more complex projects conducted in three or more locations may require a temporary shared online workspace where the lead team can post documents as they become available and update them when necessary. In addition, local teams can post questions and everyone will benefit from the lead team's answers and clarifications. Many free options are available, such as Google Groups, Ning, Nexo, Wiggio, Windows Live Groups, and Yahoo! Groups. Fee-based online collaboration sites include Basecamp, Convos, HyperOffice, ProjectSpaces, SharePoint, and Sosius. For most projects, a free service will suffice and be simple to setup. The lead team should evaluate the site's features (e.g., discussion list, calendar, file sharing, polls, task lists) and limitations (e.g., storage space, number of members) before inviting members to make sure it will meet the needs of the project. In addition, the site's security policy should be carefully reviewed to ensure that it meets confidentiality standards required for the project.

Ideally, all language versions of the materials should be available to all the teams. If something is unclear in a document translated into a local language, the team can compare their document to the one in English or another language they are familiar with, and determine the intended meaning of a term or phrase.

3.9. Briefing with Local Teams

Even if all materials seem self-explanatory, there should always be a verbal briefing between the lead research team and the local teams. If members of the lead team decide to travel to the test locations, this briefing can be conducted on site, a day or two prior to data collection. If the

timelines or budgets do not allow for travel, a phone briefing (possibly using a video conferencing tool) with the local teams will have to suffice.

In addition to some general information covered in the Test Plan and other materials, the briefing should include a detailed walkthrough of the Moderator's Guide in the context of the tested stimuli. The teams should discuss objectives for each task and question, type of data to collect, and the required degree of probing. Also, the lead team should make sure the local moderators understand task priorities and are prepared for contingencies.

When on site, a member of the lead research team should take an opportunity to talk to the interpreters if they are involved in the study. Not all interpreters are trained in the same way and should be provided with a list of guidelines, so sessions in all locations are interpreted in a consistent way. These guidelines can include translating in the first person, translating verbatim rather than paraphrasing, and translating all task instructions even though they are the same for each participant. If the lead team is not on site and it is impractical to hold an additional briefing call with the interpreters, the lead team should provide the local teams with the guidelines to be shared verbally with the interpreters before the study.

3.10. Local Pilot Testing

3.10.1. Practice and Refinement

Conducting a pilot test will allow the lead research team to check and correct the execution of the procedure prior to data collection with actual participants. If a member of the lead team is on site, the pilot session can even be conducted on the same day as the testing, as long as there is time following the pilot for any necessary course correction. If no one from the lead team is travelling to the test locations, each local team should share a video of their pilot session so that

the lead team may provide feedback. Ideally, there should be as many pilot sessions as there are moderators in the study.

Issues noticed during pilot sessions can be corrected only after their causes have been determined. Issues can arise due to an improper translation of the Moderator's Guide, the moderator not following the guide, or the interpreter's loose or incorrect translation. Thus, sometimes a change to the guide translation will be necessary, sometimes the moderator will have to be asked to follow the guide more closely, and sometimes the interpreter will need to learn new terminology, especially if it is very domain-specific.

The lead team should make sure that all moderators understand the concepts and objectives of every task and do not "just read the script." This is especially important in qualitative studies. If a participant's answer to a question in the guide is unclear or incomplete, a moderator who understands the study will probe in alternate ways, making sure he or she obtains sufficient information to satisfy the test objectives.

If it is necessary to reset the stimuli at the end of each session (e.g., delete all added contacts and appointments on a mobile device), the local team should run through this procedure at the end of the pilot session to ensure that the instructions are clear and that they work for the local version of the stimulus. Similarly, if the local test team has a dedicated note taker, he or she should use the provided datasheet during the pilot session, so the lead team can review the entries and offer feedback.

3.10.2. Internal vs. External Pilot

There are two types of pilot participants – internal and external. The internal pilot participant can be anyone from the local team who is not involved in the study (e.g., office assistant). If the

study will have two parallel sessions and there are two moderators and two interpreters, one of the interpreters can be a pilot participant and the other can interpret the pilot session. A second pilot session can be conducted with the other moderator and the other interpreter as a participant. Being a participant can be beneficial for the interpreters, especially if they are unfamiliar with the tested technology.

Internal pilot participants are usually flexible in terms of time and they do not need to be compensated. Therefore, if a member of the lead team is on site, he or she can bring up moderation and interpretation issues as they arise during the session, which is usually more effective than bringing them all up afterwards. However, this will make the internal pilot session longer than the scheduled session, which must be accommodated when creating the schedule.

An external pilot session involves a participant recruited according to the screener and who usually needs to be compensated. For a study testing a specialised product or application, an external pilot may be a better option due to the appropriate domain knowledge and experience. An external pilot session should, in general, not be interrupted, so any moderation or interpretation issues should be addressed after the session. The best approach is to conduct an internal pilot, have a debriefing with the team, and then conduct an external pilot, but this may not be feasible if budgets and timelines are tight.

Key Take-Aways

- o Identify the objectives for the study overall, as well as targeted objectives for each of the countries. Determine the reasons why certain countries were selected for the study and whether similarities or differences are the focus.

- o To avoid artifacts in the study results, review stimuli for the different test locations, make sure the right language versions are used, and, whenever possible, correct obvious localisation issues.

- o Make sure the stimuli function in all locations. The local teams should obtain the necessary hardware and test the stimuli as early as possible. Have a backup plan in case the stimuli fail to work during data collection.

- o Create a detailed Test Plan to help you communicate with the local teams and ensure consistency. The Test Plan should be a living document that includes information on the project schedule, objectives, stimuli, target user groups, methodology, lab setup, expected output, and team members' contact information and availability.

- o Create and localize the Screening Questionnaire, which defines how participants in each country will be selected. Adjust the number of recruits based on local no-show rates and set the test schedule to accommodate local holidays and customary behavior.

- o Keep consent forms and non-disclosure forms brief and simple. Rely on the local versions of these documents instead of creating your own.

- o Decide on the appropriate amount and form of participant compensation in each country.

- o Prepare a detailed and explicit Moderator's Guide. Include what the moderator will say to the participant, as well as information and instructions for the moderator. Refine the guide through pilot testing prior to localisation.

o   Localize the Moderator's Guide using the local teams or professional translators but be aware of the advantages and disadvantages involved in each of these options. A Moderator's Guide translated professionally should still be reviewed by the local teams in the context of the stimuli.

o   Consider reverse translating materials for studies requiring high precision.

o   Make all recruiting, test, and reporting materials available to the local teams ahead of time. Include a video from a session conducted in the first location of the study. For studies with three or more locations, use an online workspace to share the materials and post questions and answers.

o   Conduct a briefing prior to the study either in person or via conference call. Include a detailed walkthrough of the Moderator's Guide in the context of the stimuli.

o   Provide the interpreters with guidelines to which they should adhere when simultaneously translating the sessions.

o   Have the local teams conduct a pilot test prior to data collection, so you have a final chance to correct any translation, moderation, or interpretation issues.